MODELLING AND FORECASTING THE DEMAND FOR ELECTRIC ENERGY IN NEW ZEALAND

A thesis presented for the
degree of Master of Engineering in Electrical Engineering
in the University of Canterbury

by

M.J. Turnbull B.E.(Hons)

1974

ABSTRACT

This thesis studies the demand for electric energy and ways of forecasting it, as an aid to the economical design and operation of electric power systems. An examination of the nature of consumers demands leads to a two part model of the demand. Long term growth of demand is shown to be determined by the way the numbers of appliances owned by consumers increases, while the short term daily, weekly and seasonal demand fluctuations result from the way consumers use their appliances. A number of forecasting methods utilizing this model are studied.

The accuracy of a demand forecast influences the amount of reserve capacity needed to satisfy a given level of demand with a specified degree of reliability. A criterion is presented which determines when a forecast may be considered sufficiently accurate. Sufficient accuracy is defined in terms of minimizing the costs of providing the reserve capacity and of improving the accuracy of the forecasts.

ACKNOWLEDGEMENTS

I wish to thank the many people who have helped me, through advice, discussion, criticism and encouragement, to complete this thesis. Special thanks are due to my supervisors, Dr. J.H. Andreae and Dr. H.J. Jelinek. Also I must thank the New Zealand Electricity Department for the use of their data and the Royal New Zealand Air Force for the leave of absence to carry out this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 A Background to the Thesis

A comprehensive model of the load on the New Zealand electric power system is the long term goal of the work presented in this thesis. This model is to be used for forecasting consumer demands for electric energy.

Electric energy cannot be stored economically in the amounts typical of a power system and must be generated as required. To maintain the quality of the energy supply (i.e. voltage and frequency) at predetermined levels the power output of the generating system must remain equal to the demand for energy. Output cannot respond instantaneously to changes in demand due to restrictions imposed by plant parameters (3, 4, 5) or, in the worst case, because the total generating capacity of the system is exceeded. Demand forecasts enable the initiation of adjustments to the system's power output before any changes in the demand occur. The lead time of the forecast should equal, or exceed, the time required for the particular adjustment.

System operation requires demand forecasts over lead times from a few minutes to a few days for

(a) the adjustment of the output of plant already operating, and

(b) scheduling the start up of additional generators.

Forecasts of both demand and energy requirements over lead times of several years are used by system planners for scheduling the installation of additional plant. Thus future system operators must operate the system to meet the actual demands from the amount of generating capacity installed on the basis of these long term forecasts.

The management of a power system may elect to meet all demands for electric energy from consumers or, alternatively, devise controls to restrict the demand to some desirable value. For example, supply authorities in New Zealand control electric water heating with the aim of reducing costs to the consumer. The procedure for designing appropriate controls is equivalent to forecasting the unconstrained demand and then determining the adjustments required (6). In this situation a model of the load is required which is able to confirm that a proposed control action will have the desired effect.

Several methods of forecasting demand and energy requirements over all lead times have been described in the literature. A comprehensive and critical review of these methods is contained in chapter 4 of this thesis; shorter reviews are given by Stanton and Gupta (1) and Matthewman and Nicholson (2). All methods require a model of the load and its behaviour; this may be a time series representation of historical demands or a more complex representation of the structure (or rules of behaviour) of the load. The best forecasts, in a minimum error sense, can only be obtained if the model is appropriate to the particular load (10, 11).

The New Zealand power system is composed of a central (state-run) generating organization and 69 autonomous supply authorities (7, 8).

The central organization generates and sells electric energy in bulk to the supply authorities which in turn distribute it to individual consumers. At present supply authorities estimate their annual maximum demands and energy requirements in isolation. These separate forecasts are then combined to form national forecasts which are then used in planning future plant requirements (7, 8, 9). In this way specialist local knowledge is introduced into the national forecasts; without it there are limits on the detail in the national forecasts (11, 17). The literature contains no evidence of any study of the New Zealand load to ensure the models in use are appropriate. Such a study forms a considerable part of the work presented here. This thesis brings together, for the first time in New Zealand, a comprehensive review of available forecasting methods and a comprehensive study of the New Zealand load.

1.2 Definition of Problems Examined in the Thesis

An "overview" of the place of a comprehensive load model in power system planning and operation is provided in Figure 1.1. The actual demand for electric energy results from the action of the environment in which the power system exists on the consumers. In general there is a time lag between the occurrence of an "event" in the environment and a change in the demand in response to that "event"; this time lag varies between environment variables. If a known environment is applied to an appropriate model of the load then the demand can be forecast over lead times at least equal to this lag. Over longer lead times a projected environment must be used. The application of an actual or projected environment to the load model, resulting in a forecast value of demand, is shown in parallel with the real world in Figure 1.1. Information on actual and forecast demands is supplied to the generation planning and decision processes. These processes determine the amount of generation capacity (including reserve) required to ensure the forecast demand will be met with the desired degree of confidence (usually of the order of 99.99% (16)); the allocation of reserve capacity is discussed in greater depth in Chapter 3.

Two particular aspects of figure 1.1 are examined in this thesis;

(a) the modelling of the load and its response to the environment, and

(b) the accuracy with which it is necessary to forecast the demand.

The way in which the environment determines the demand is shown in greater detail in Figure 1.2. Not all environment variables have the same effect on the load; in Figure 1.2 these variables form two distinct groups;

(a) those which govern the number of appliances which consumers have available for connecting to the distribution network (i.e. the potential demand), and

(b) those which determine the way consumers use the appliances which they own.

For the majority of appliance types only a few discrete rates of energy usage are possible. These are fixed when the appliance is designed. In an ideal case, a knowledge of the number of appliances, their types, and the way they are used would be sufficient to determine the demand.

The problems of long and short term forecasting are illustrated by Figure 1.2. Short term forecasting (lead times up to a few days) is essentially concerned with how consumers use their appliances as, over the usual lead times, the appliance numbers are known and relatively constant. Provided the way appliances are used remains substantially constant over periods of several years the long term forecasting problem is essentially the determination of growth in appliance numbers.

To summarize, the principal problems examined in this thesis are

(a) the modelling of load growth as represented by the growth of appliance numbers,

(b) the modelling of the way these appliances are used by consumers, and

(c) the accuracy with which it is desirable to forecast the demand (and hence how well the models must represent the real system).

1.3 Thesis Organization and Chapter Summary

Each chapter of the thesis is as self-contained as possible, consistent with a satisfactory treatment of the problems. Later chapters build on earlier work with a minimum of references in earlier chapters to later ones. The basic structure of the thesis is shown in Figure 1.3. There are two parallel lines of development. In Chapter 2 the overall load modelling problem is formulated; this leads directly to the discussions of load growth in Chapter 5 and of the demand (or way appliances are used) in Chapter 6. In parallel with this modelling theme is the discussion of forecasting accuracy in Chapter 3. This is followed by a review of the state of the art (i.e. load modelling and demand forecasting) in Chapter 4. The approaches adopted in Chapters 5 and 6 are influenced by this review. Chapter 7 summarizes the principal achievements of the thesis and makes recommendations for further work. Appendices are used for derivations, experimental results and a certain amount of data which is presented as a sequence of diagrams. References are listed together at the end of the thesis, in order of appearance.

Detailed formulation of the modelling problem is the subject of Chapter 2. This formulation, following the outline in Figure 1.2, treats the problem in two parts. The first part is modelling the process governing appliance number growth. Modelling the usage of appliances forms the second part.

In Chapter 5 the load growth process is further divided into two parts:

(a) that which determines consumer numbers, and

(b) that which determines the number of appliances a consumer owns.

The nature of these processes varies between consumer types. A type of birth and death process is advanced to explain domestic consumer numbers. Using this model consumer numbers can be forecast for lead times up to about 15 years from known information. This is superior to models which require the prior extrapolation of "independent" variables, for example the "housing starts" index in the correlation model used by Godard (12). Electrical appliances are treated as a sub-class of the class of goods known in economics as consumer durables. The work of Pyatt (13) on models of the accumulation of durables by households is used to develop a model of the growth of a domestic consumer's ownership of appliances.

Non-domestic consumer numbers cannot be modelled by a simple birth and death process as in the domestic case. Neither are there any useful (for forecasting) rules governing the number of appliances owned by individual consumers.

Models of appliance usage, which are the subject of Chapter 6, have been based on day to day changes in the daily load curves. These changes reflect day to day variations in the way appliances are used. The shape of the daily load curve itself is determined by the aggregation of these usage patterns over the day. While the model development actually proceeded in an iterative fashion the chief results only are presented; a complete record would occupy too much space and serve no useful purpose.

In the approach adopted here each day of the week is modelled separately, thus avoiding the need for the day-of-week factors etc. used, for example, by Davies (14); this approach is equivalent to the weekly load curve approach used recently by Christiaanse (10). In the results and discussion attention is confined to one particular day of the week - Tuesday.

The discussion starts with a model of the observed seasonal changes in daily load curve shape. Three daily load components, the shapes of which are obtained artificially, are shown to be sufficient to explain the mean seasonal behaviour. With this model it is possible to generate a sequence of daily load curves for a complete seasonal cycle knowing only the average and expected maximum (or minimum) daily energy usage for that seasonal cycle. The seasonal movements are assumed basically sinusoidal in nature.

A Fourier series representation of the mean seasonal behaviour is also discussed. It is shown that an adequate representation of the mean behaviour is obtained using only a constant term and the first sine term to represent the integrated demand in each daily interval. The Karhunen-Loeve expansion (2,15) is also investigated as a model of seasonal behaviour; it turns out that it is not suitable for this purpose although it has been used in very short term forecasting applications (2).

In a third model demand is related to the internal temperature of buildings and to hours of daylight. A complex non-linear relationship between demand and internal temperature was first used by Davies (14) to model the demand at the time of the daily maxima. In Chapter 6 a piece-wise linear relationship is applied to all intervals of the day. The presence, or absence, of daylight at each daily interval (which is a function of the time of the year) determines whether a constant block of load, attributable to lighting, is present or not.

A considerable amount of variation in the demand is not explained by this model. Neither can it be explained satisfactorily by introducing further weather parameters. It is concluded that this variation arises from vagaries of human behaviour and that it cannot be reduced without extremely comprehensive (and expensive) load control measures. This residual variation sets an upper limit on the achievable demand forecasting accuracy. Considerable improvement in weather forecasting accuracy is needed if this upper limit is to be achieved in practice.

In Chapter 3 a criterion is derived for determining the accuracy with which it is desirable to forecast the demand. The desirable accuracy minimizes the total cost of allowing for uncertainty in the forecast and of making the forecast. Using this criterion it is shown that there is no advantage in reducing the uncertainty in demand (and energy) forecasts below an amount set by plant unreliability.

This criterion also means that forecasting methods must be compared on the basis of the uncertainty in the forecasts at some stated confidence level; the amount of spare capacity required to ensure the supply achieves the reliability specified by management is directly related to this uncertainty. The commonly used measure of comparison between methods is the achieved forecasting error (i.e. the difference between forecast value and actual value) (1, 11, 17); this may lead to misplaced optimism as to the forecasting accuracy of a method and to insufficient reserve capacity being provided for a given reliability requirement.

A comprehensive review of existing electric energy forecasting methods constitutes Chapter 4. Methods are compared using the uncertainty measure derived in Chapter 3. The methodology of demand and energy forecasting over all lead times is examined critically.

1.4 Data Processing and Computer Programs

The considerable amount of data processing required during this project was carried out with the aid of the computer at the University of Canterbury (an IBM 360/44). A number of computer programs were written to perform specific jobs, for example, the evaluations of the trend extrapolation model, the scaling metho and the Karhunen-Loeve expansion method in Chapter 4 and the regression analysis in Chapter 6. For simplicity these programs employed standard sub-routines, where these were available, to carry out the required computations, e.g. statistical analysis, regression analysis, eigen value evaluation and various matrix operations such as inversion. Most programming effort was devoted to organizing data for input to these sub-routines and to displaying the results of the computations on the line printer and on an incremental drum plotter. No new computational techniques were developed as such.

The majority of the data used and some programs relating to its input to the computer are described in a Departmental Memorandum, see ref. 18. Several other sub-routines*, for displaying data on the line printer or drum plotter, and assembly language versions (for increased speed) of standard routines are available in the Electrical Engineering Department Program Library. A number of illustrations in this thesis (e.g. Figure 2.2) were prepared using one of these Fortran routines which plots a surface as sequence of cross-sections in the Z-Y plane. The specific nature of the remainder of the programs meant that they would be of no benefit to others. Sufficient information is given in the text to enable a reasonably competent person to program any of the computations described.

* written in Fortran IV language for the 360/44 computer.

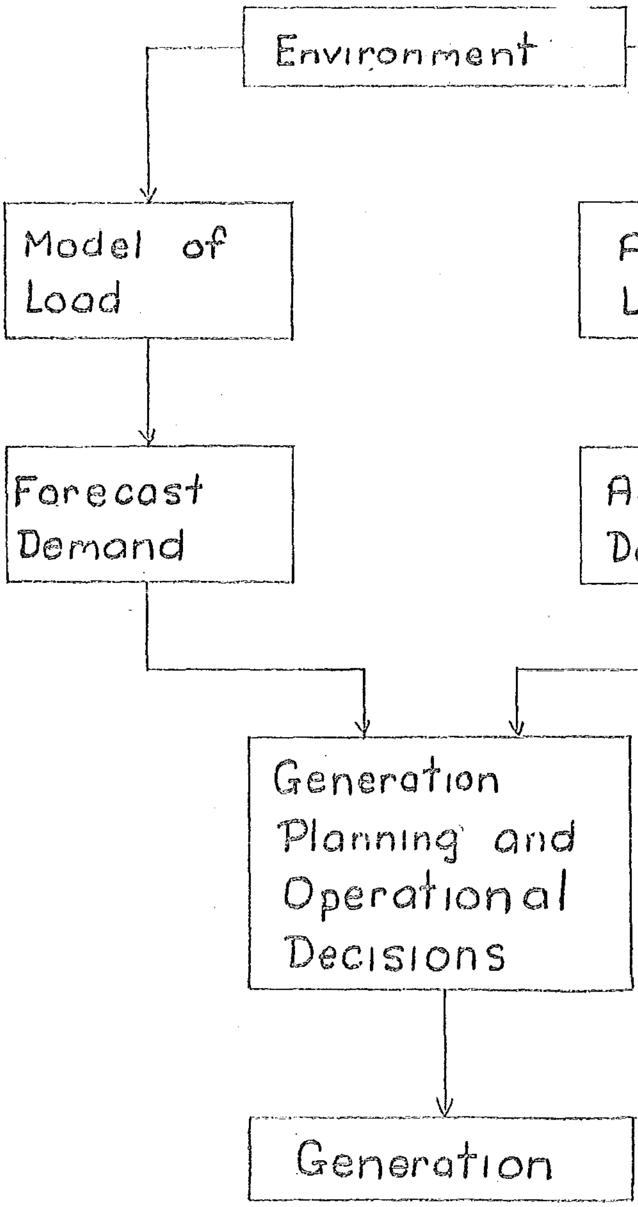


Figure 1.1 The place of a model of the load in system operation and planning.

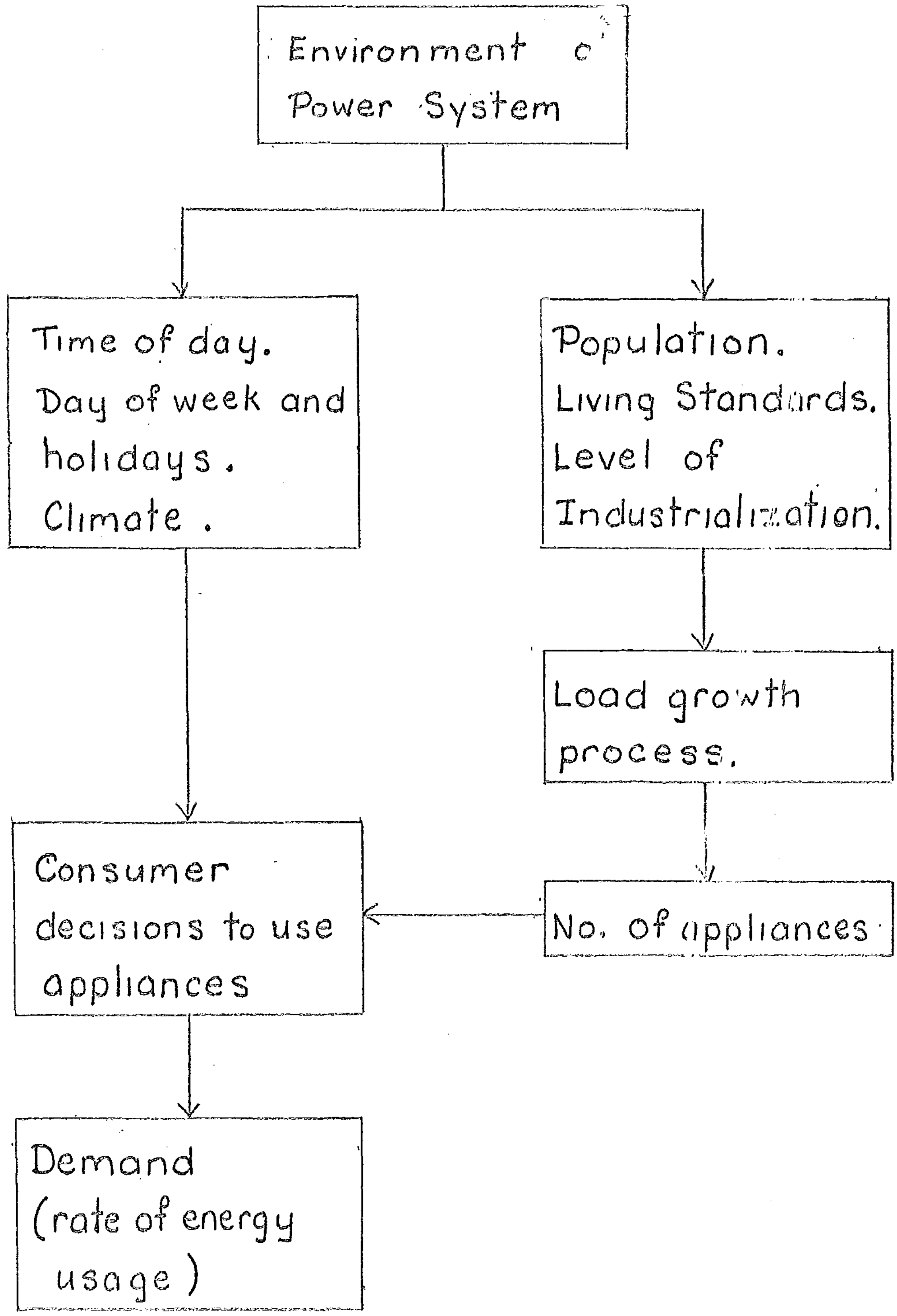


Figure 1.2 Influence of the environment on the demand for electric energy.

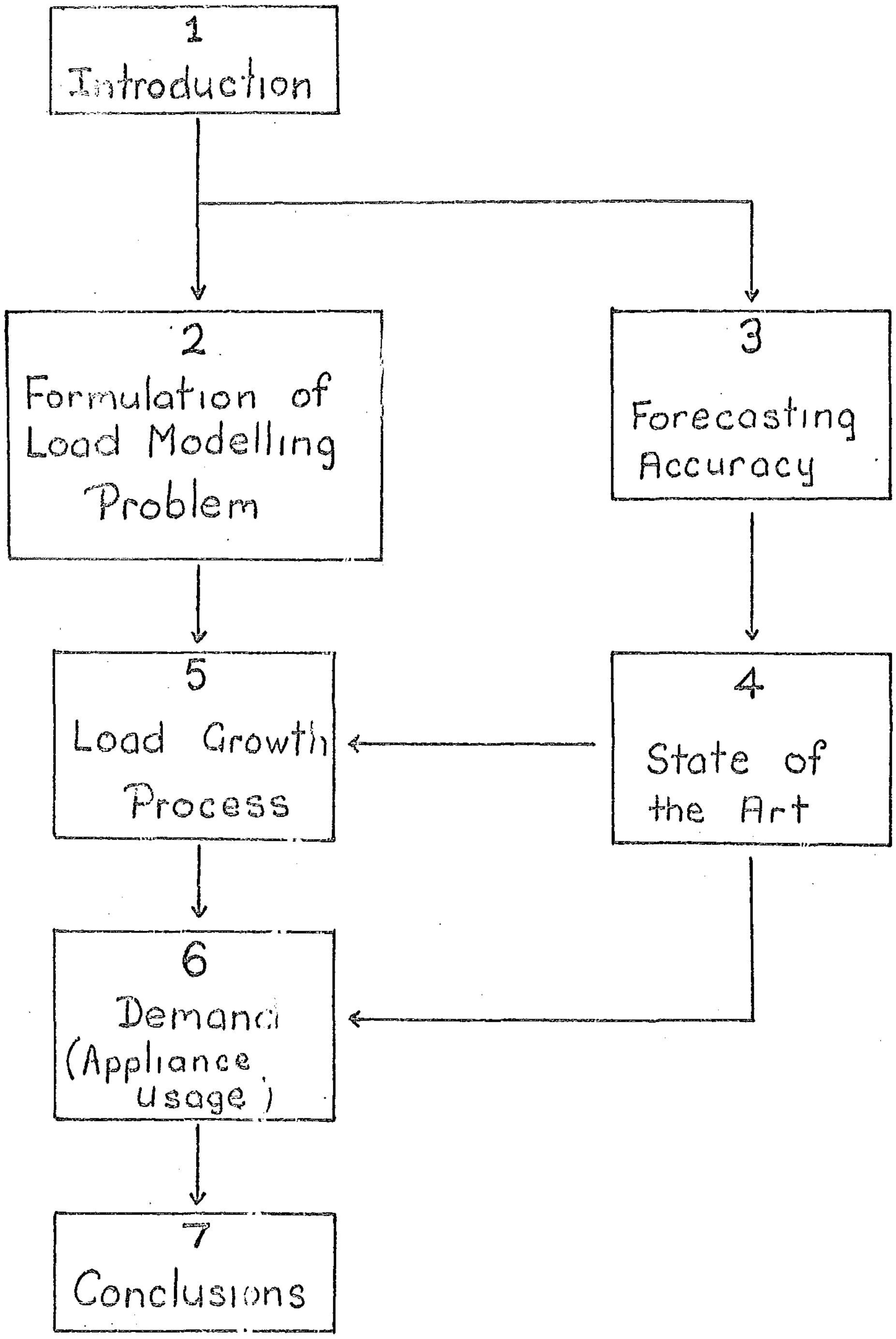


Figure 1.3 Thesis structure

CHAPTER 2

THE NATURE OF THE DEMAND

2.1 Introduction

An electric power system supplies electric energy to consumers distributed throughout a geographical region which may be a city, borough or county. This chapter discusses the consumers and their uses for electric energy. The essential features of their demands for electric energy are determined. These features form the basis for the models of the short and long term time behaviour of the demand which are developed in later chapters. The discussion is illustrated by the historical behaviour of the demand in New Zealand.

2.2 Classification of Electric Energy Consumers

On the basis of their primary use for electric energy the consumers within the region may be grouped into four general classes. These classes are:

(1) Domestic consumers

A domestic consumer is a group of people, e.g. a family, living together in the same dwelling; c.f. the household of economic theory (13, 19, 20).

(2) Industrial consumers

An industrial consumer is an establishment, e.g. a factory, which manufactures, mines or processes a particular good or range of goods for consumption within, or export from, the region. Energy, from one or more sources, forms an input to the manufacturing process, together with labour and raw materials and is used for both work, i.e. motive power, and heat. The energy requirements of each consumer are related to the amount of output, by some function of the particular process.

(3) Commercial consumers

A commercial consumer is an establishment trading in merchandise or professional services etc. Such establishments include shops, offices, warehouses etc. Energy is used for heating, lighting and other services related to the premises, e.g. escalators. Energy is not, in general, essential to the operations of these consumers but is used for staff and customer convenience.

(4) Public services

These consumers provide services to the public such as street lighting, transport (but not including fossil fuelled vehicles), sewage disposal etc. These are, in New Zealand, considered socially necessary and are commonly operated by bodies such as city councils. Energy requirements are related to the number of people served; e.g. street lighting requirements are a function of urban road length which is in turn some function of the number of urban households.

These classes may be further subdivided in particular cases, e.g. Jelinek (21) has subdivided class 1 on the basis of income and whether the dwelling is a flat or house etc.

The number of consumers in each class increases with time (as some function of the increase in population) and the ratio of consumer numbers between classes also changes with time. This is illustrated in figure 2.1 for domestic and non-domestic consumer numbers in New Zealand.

2.3 The demand for electric energy

2.3.1 Some Terminology

Consumers in all classes request a supply of electric energy by connecting appliances to the supply network. These connected appliances form the LOAD on the power system. The DEMAND for electric energy is the rate at which the load uses electric energy; it is usually measured in units of power, i.e. kilowatts (kW). Demand may be referred to as -

(a) instantaneous demand which is the rate of energy usage at an instant in time, or

(b) integrated demand which is the average rate of energy usage over a time interval of arbitrary length; usually one quarter, one half or one hour. (22, 23).

The two terms are related by

$$D_i = \frac{1}{\tau} \int_{T_i}^{T_i + \tau} D(t)\, dt \qquad (2.1)$$

where τ = length of interval in hours

$D(t)$ = instantaneous demand at time t, and

D_i = integrated demand over the i th interval.

The total amount of energy used by the load over a period $(0, T)$ hours is, using instantaneous demand

$$W = \int_0^T D(t)\, dt \qquad (2.2)$$

or using integrated demands

$$W = \tau \sum_{i=1}^{n} D_i \qquad (2.3)$$

where n = an integer $\geqslant 1$ denoting the number of intervals of length τ hours in the period $(0, T)$

2.3.2 The variation of demand with time

Each connected appliance contributes to the demand. The total demand within the region at a particular time is the sum of the demands of all the appliances connected at that time. Define A to be the set of all appliances owned by consumers and able to be connected at any time. Not all appliances are necessarily connected at time t ; let α_t denote a set made up of members of A which are actually connected at time t : $\alpha_t \subseteq A$. If j denotes a member of A the total instantaneous demand at time t may be written as

$$D(t) = \sum_{j \in \alpha_t} c_j(t) \tag{2.4}$$

where $c_j(t)$ = demand of member j at time t.

The magnitude of the demand varies in a cyclic manner with time of day. The demand at a particular time of the day also varies with the day of week and time of year (2, 24). In figure 2.2 the shape of the daily load cycle for different days of the week and for consecutive weekday average load curves through the year for an urban supply authority are shown.

2.3.3 The number and type of the connected appliances

Any member of the set A may, with a few exceptions, be connected or disconnected at any time. A single appliance can be either 'on', i.e. connected, or 'off', i.e. disconnected, at an instant in time (22, 25). Let M denote the number of different types of appliances in A, and $m_i, i = 1, .., M$, the number of appliances of each type. Assume independent use of appliances of each type; this means that whether an appliance is used or not is a function of the appliance type but not of the consumer owning it. The probability that a type i appliance is 'on' at time t is $p_i(t)$; the corresponding 'off' probability is $q_i(t)$. Provided that the probability of observing an on-off or off-on transition is negligible then

$$p_i(t) + q_i(t) = 1 \tag{2.5}$$

The probability that there will be exactly n_i appliances of type i in the set α_t is given by the n_ith term of the expansion of

$$(p_i(t) + q_i(t))^{m_i}$$

i.e. p (α_t contains n_i type i appliances) =

$$m_i C n_i \quad (p_i(t))^{n_i} \quad (q_i(t))^{m_i - n_i} \tag{2.6}$$

The mean and variance of the distribution of possible numbers of type i appliances on at time t is given by

$$\begin{aligned} \bar{n}_i &= m_i p_i (t) \\ \text{var}(n_i) &= m_i p_i (t)\, q_i (t) \end{aligned} \tag{2.7}$$

The contribution of type i appliances to the total demand is obtained by weighting the number of connected appliances (equations 2.7) by the demand of this type of appliance. For single element appliances, e.g. lights, refrigerators, irons, this is merely the nameplate rating. For multiple element appliances, e.g. electric ranges, radiators etc., it is necessary to know which elements are connected at the time, and this is likely to be different for each example of the type. Once the elements are selected by the consumer the demand remains constant with time. Some industrial loads, e.g. motors, are exceptions as their demand varies in time as a function of the load on the motor. However for these and multiple element appliances the expected demand when connected may be used (25). The mean and variance of the demand from type i appliances becomes

$$\overline{D_i}(t) = m_i \bar{c}_i p_i(t)$$

$$\mathrm{var}(D_i(t)) = m_i \bar{c}_i^2 \, p_i(t) \, q_i(t) \qquad (2.8)$$

where $\bar{c}_i$ = expected demand of type i appliance,

i = 1, ..., M

The total demand from all types of appliance at time t is

$$D(t) = \sum_{i=1}^{M} D_i(t) \qquad (2.9)$$

The form of the distribution of D(t) is obtained by combining the M distributions of equation 2.8. Provided the demands $D_i(t)$ are independent and M is sufficiently large the central limit theorem may be applied (25, 26). The distribution of D(t) is then approximately normal with mean and variance given by

$$\overline{D(t)} = \sum_{i=1}^{M} m_i \bar{c}_i p_i(t) \qquad (2.10)$$

$$\mathrm{var}(D(t)) = \sum_{i=1}^{M} m_i \bar{c}_i^2 \, p_i(t) \, q_i(t)$$

If the set A does not change with time, i.e. M and m_i remain constant - which means no growth in appliance numbers, and if the $\bar{c}_i$, $i = 1,..,M$, do not change between consecutive periods of use then the only way for the demand to vary is for the $p_i(t)$ to vary with time.

Hence the time sequence of $p_i(t)$ has the same shape as the time pattern of use of appliances of type i. The shape of the load curve is formed from the combination of these appliance usage curves.

2.3.4 The determinants of appliance usage patterns

The surroundings and circumstances in which a consumer exists, i.e. the environment, determine the way in which a consumer uses its appliances. The social, economic and climatic aspects of the environment appear to have the most influence.

The social environment governs the living and working habits of the regional population. It determines, for example, the times of the day when people eat their meals, watch television programmes and go to bed (21, 24, 27). It also dictates that the majority of industrial and commercial consumers do not operate during weekends; thus the on-probabilities of these consumers' appliances is a function of the day of the week. The social environment dictates that the probability of an appliance being on is a function of the time of the day.

The economic environment affects the level of the demand for goods and services within the region (19, 28, 12). An increase in this level, for a fixed population, requires an increase in the amount of production (28). Initially this increased production is obtained by working plant for longer hours, thus extending the time for which the on-probability for industrial appliances is high. Thus this aspect of the environment modifies the basic time-pattern of use determined by the social environment. There is no evidence that domestic appliance use is affected by the economic environment although it affects the number of appliances a consumer owns (13, 28, 29).

The climatic environment directly affects consumers' heating and lighting requirements (14, 30). The seasonal decreases in mean air temperature, for example, increase the probability that heating devices will be in use. Similarly the increase in hours of darkness increases the period of time for which lighting is in use, thus modifying the time-pattern of use of lighting appliances.

The social environment dictates that the probability of an appliance being on is a function of the time of the day; the other aspects of the environment modify the basic time patterns. Let $\tilde{e}_i$ denote the set of environment variables relevant to the use of type i appliances. The influence of the environment on D(t) may be incorporated into equation 2.10 by treating t as the time of day and modifying $p_i(t)$; i.e.

$$p\,(\text{type } i \text{ appliance is on}) = g_i(\tilde{e}_i\,,\, t) \qquad (2.11)$$

The mean and variance of the distribution of D(t) then become

$$\begin{aligned} \overline{D}\,(t) &= \sum_{i=1}^{M} m_i c_i g_i\,(\tilde{e}_i\,,\, t) \\ \text{var}\,(D(t)) &= \sum_{i=1}^{M} m_i \bar{c}_i^2 g_i(\tilde{e}_i, t)(1-g_i(\tilde{e}_i, t)) \end{aligned} \qquad (2.12)$$

The extent to which any single environment variable affects the shape of the daily load curve is determined by the number of appliances it can affect. In particular, where a load has a large industrial or public service segment the influence of the climate is greatly reduced. This effect is illustrated in figure 2.3 which shows monthly average daily load curves (weekdays only) against time of year for three NZED loads. The first was mainly domestic with no large industrial consumers, the third was mainly domestic but with one large industry and the second included railway traction and a large public lighting load (a road tunnel).

2.4 Growth in Energy Usage

2.4.1 The demand for energy from all sources

To the consumers electricity is merely a source of energy, though with some unique features such as ease of control and cleanliness in use. Alternative sources which a consumer may choose for a particular application include coal, oil and gas. The factors which a consumer considers when making the choice may be summarized as (13, 28, 31, 32):

(a) relative cost of the alternatives,

(b) availability and ease of use of alternatives,

(c) relative efficiencies, and

(d) consumer tastes and preferences (which may be influenced by advertising and social pressures (27, 32)).

This list is not exhaustive and the weighting given each factor varies between different consumers. As a consequence of this competition between energy sources electricity supplies only a portion of the regional energy requirements.

The percentages of New Zealand's total energy requirements met by coal, gas, electricity and oil for the years 1950-1967 are shown in figure 2.4. In some applications there is only one commercially feasible source, e.g. oil for motor vehicles. Thus the considerable increase in motor vehicle numbers in this period has contributed disproportionately to the oil share; in 1967 motor vehicles used nearly half of the oil (29).

2.4.2 Historical growth of electric energy usage

The amount of electric energy used annually in New Zealand is increasing approximately exponentially at a rate of about 7.5% per annum as shown in figure 2.5. This figure also shows the annual half hourly maximum demand is growing in a similar way at a similar rate.

In the period covered by figure 2.5 the population has also grown but at a lower rate; approximately 2% per annum (33). An increase in the number of households (domestic consumers) has accompanied the population increase. This is shown in figure 2.1. The average number of people per household has decreased in this time reflecting tendencies for young people to leave home at an earlier age and for young married couples to form households of their own rather than live in the family home (33).

This increase in population has increased the demand for consumer goods and services. Together with a steady increase in the level of industrial activity this has meant an increase in non-domestic consumer numbers, see figure 2.1. The rate of increase of these consumer numbers has been greater than that of domestic consumer numbers as a result of increased emphasis on an industrial base for the national economy (29, 33).

These increases in consumer numbers cannot alone explain the observed increases in annual energy usage. This means that the average energy usage per consumer has also increased in this period, as shown in figure 2.7. Three hypotheses can be advanced to explain the increases in average energy usage:

(a) the annual energy usage of newly formed consumers exceeds that of consumers already in existance, which remains constant. This appears to apply particularly to factories etc where it is difficult to increase energy usage without extensive changes to plant and working hours (29);

(b) there is an increase in consumers' usage of existing appliances. This occurs if more overtime, for example, is worked. In the domestic sector there is increasing reliance on electricity rather than coal fires for heating the home; e.g. in the 1961 census 9.9% of households claimed "mainly electric heating" and this figure increased to 38.6% in the 1966 census (33). This means that existing electric radiators etc are likely to get more use.

(c) There is an increase in the number of appliances owned and used by consumers. In Table 2.1 the numbers of various domestic appliances sold during the years 1958-1967 are shown. This data shows a marked increase in the number of electric radiators sold about the years 1961-1962. Much of this increase reflects the increased reliance on electricity for home heating observed in the census figures. Furthermore there has been an increase in the percentage of households with all electric cooking as the following census figures show (34):

1956 - 56.9%; 1961 - 68.8%; 1966 -78.6%

Annual sales of all the appliances in Table 2.1 are roughly twice the increase in domestic consumer numbers which suggests many sales are for replacement purposes and, or, many households own several appliances of one type.

The above discussion indicates that all three hypotheses are true, at least in part. Unfortunately there is a lack of data to confirm them satisfactorily (29). It is clear, however, that two processes contribute to the growth in annual energy usage:

(a) the increase in population and level of industrial activity and

(b) the increase in the average energy usage of consumers, which arises from the accumulation of appliances by domestic consumers and from increases in the size of new industrial plant.

2.5 Summary

The demand of the system load has been shown to be the sum of the demands of all the connected appliances. Of the set A of all appliances owned by consumers only a subset α_t is connected at any time t. The variation of demand with time reflects changes in both the set A and the subset α_t.

For short periods of time, e.g. one or two months, the set A may be assumed constant. The shape of the daily load curve is then determined by the daily time sequence of probabilities that different types of appliances are connected. These probabilities are functions of the time of day and the environment in which the consumer exists.

Over longer periods of time, e.g. six months to several years, consumers purchase additional appliances. Increases in the regional population and the demand for goods and services within the region result in increases in the number of consumers, both domestic and non-domestic. Both these factors increase the number and type of members of the set A. Consequently, the level of demand at corresponding times in consecutive years increases as does the amount of energy used annually.

A comprehensive model of the demand for electric energy thus has two parts. The first models the process which chooses the subset α_t of the set of all appliances and hence the short term variation of demand. The second models the process which increases the number and type of appliances which consumers own and hence the growth of demand and energy usage.

The next chapter examines the faithfulness with which these processes need be modelled.

TABLE 2.1

Annual sales of consumer appliances (000's)

Type	Year										
	1958	'59	'60	'61	'62	'63	'64	'65	'66	'67	'68
Electric ranges	29	35	33	38	42	38	40	44	50	52	47
Refrigerators	63	51	50	56	44	38	54	53	61	73	73
Radiators	35	51	47	89	155	136	152	218	222	213	220
Washing machines	37	36	38	48	40	40	42	46	47	51	44
Toasters	29	34	32	52	64	44	62	83	67	64	75
Irons	9	20	35	59	53	43	78	82	67	63	67

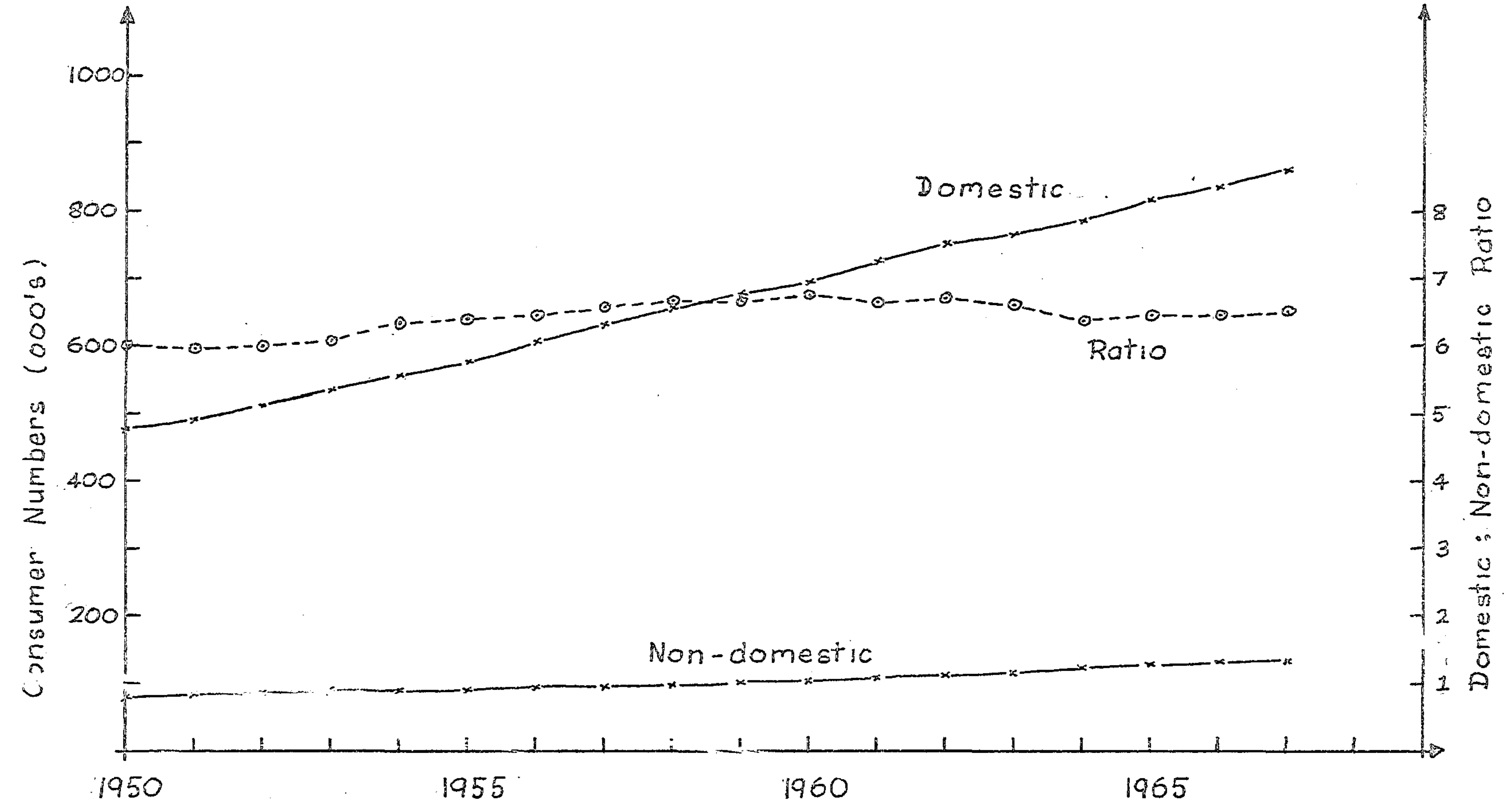


Figure 2.1 Number of Domestic and Non-domestic Consumers and their Ratio: New Zealand

ELEC 052

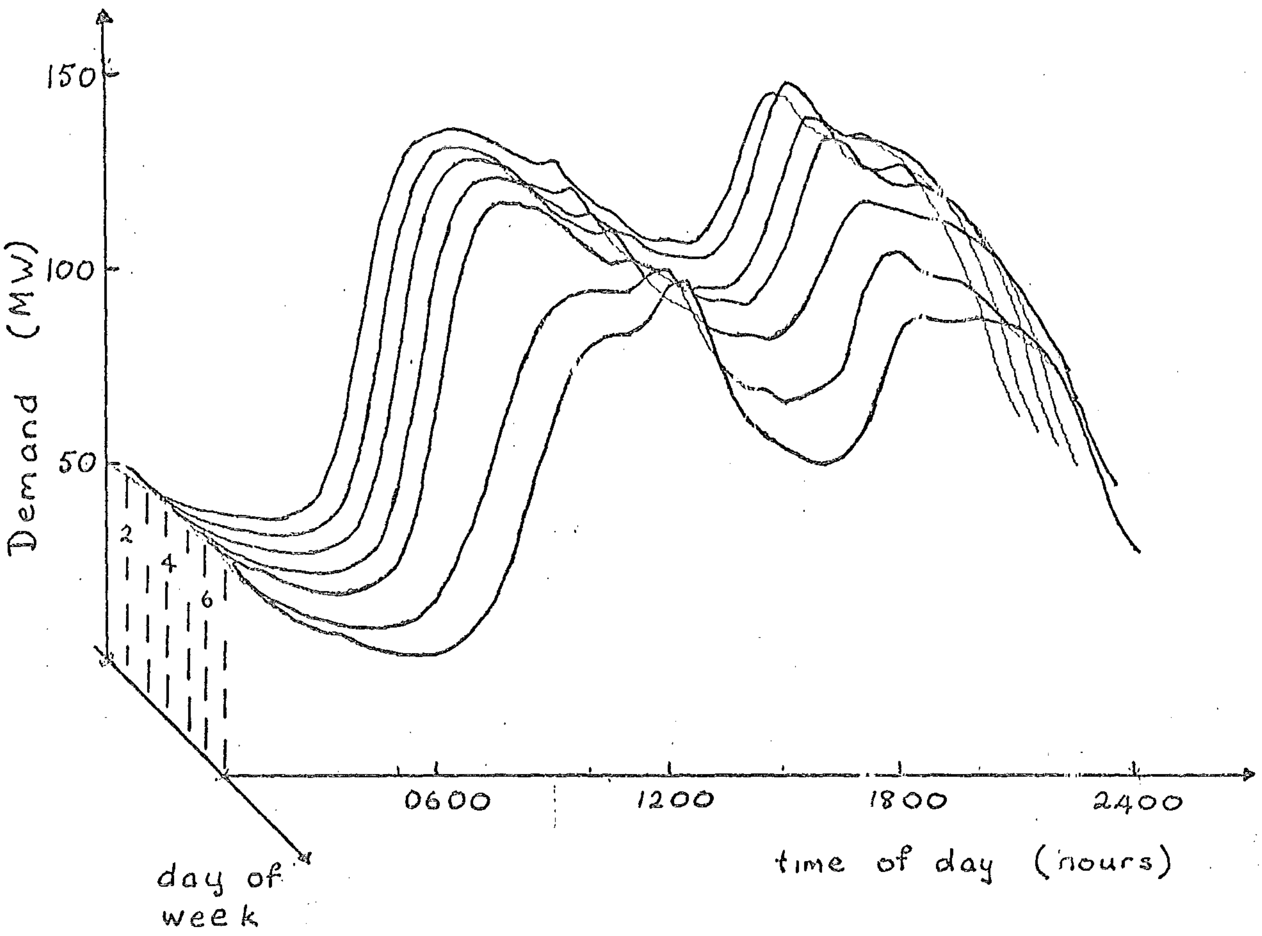


Figure 2.2 (a) Daily load curves for each day of the week (annual average) ; Christchurch MED.

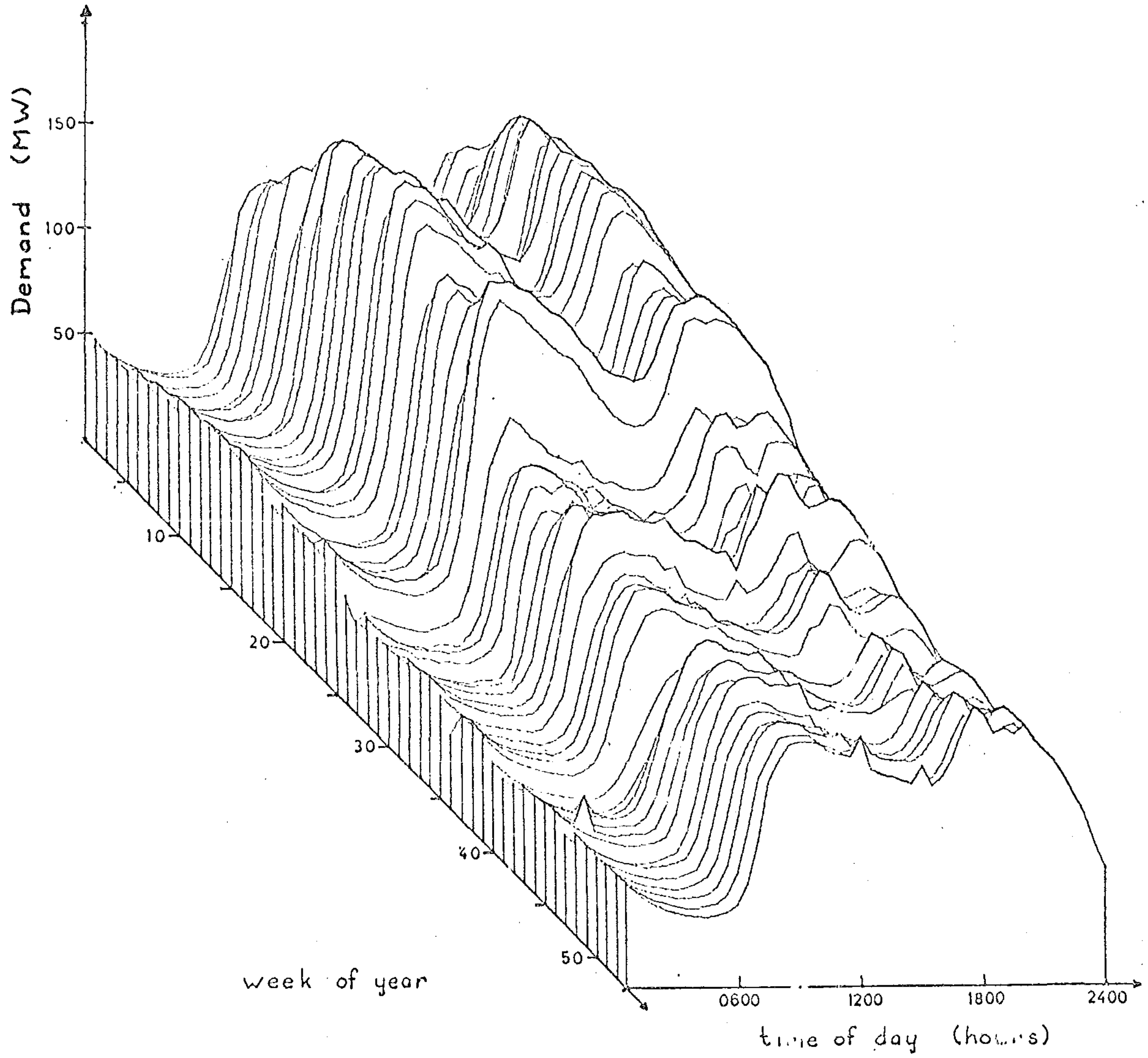


Figure 2.2 (b) Mean weekday load curves for year from 27.3.67 to 31.3.68 : Christchurch MED.

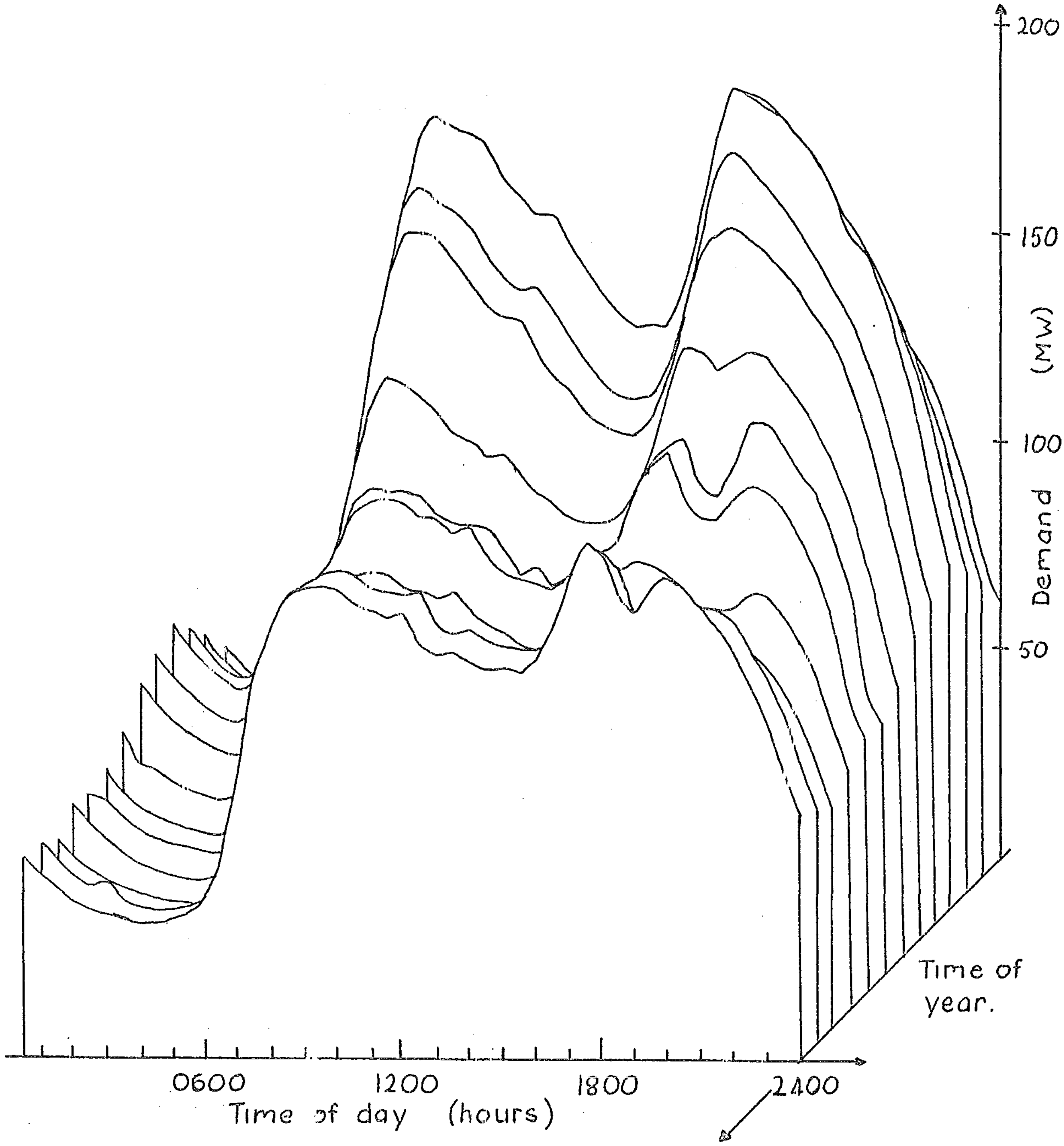


Figure 2.3(a) Monthly average daily load curves; Christchurch MED, April 1967 - March 1968

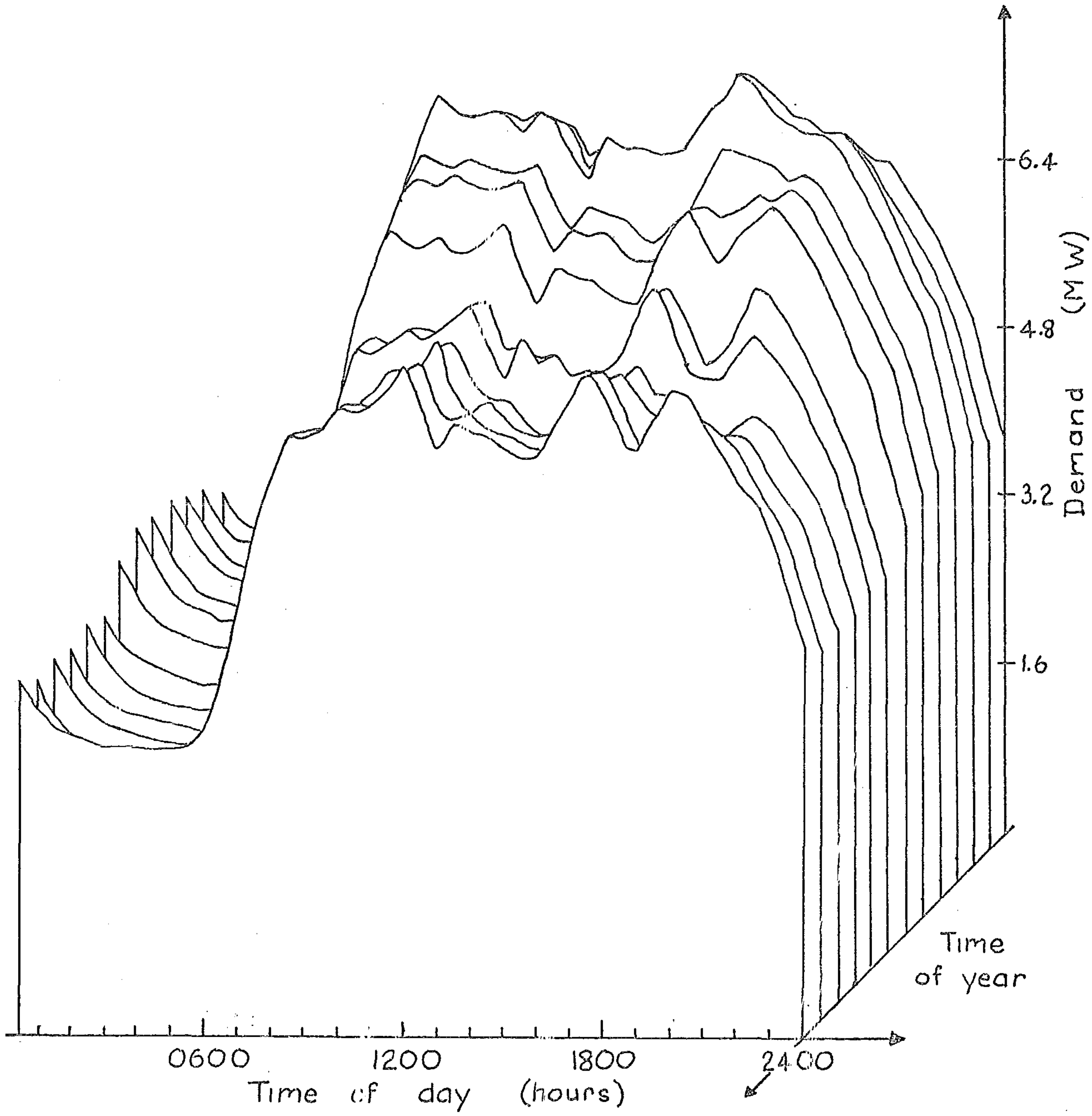


Figure 2.3 (b) Monthly average daily load curves for Woolston 1 & 2 feeders, April 1967 - March 1968.

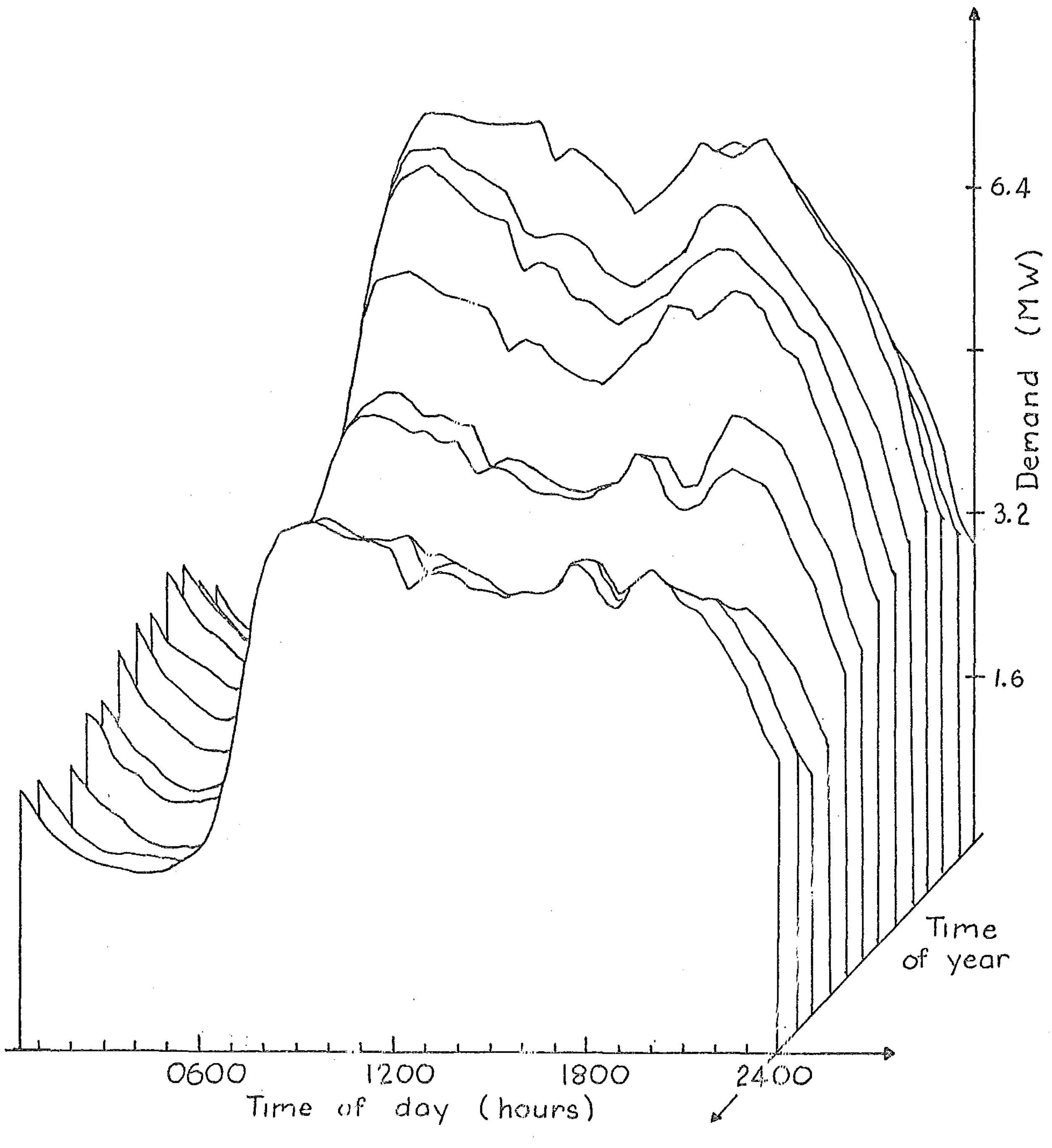


Figure 2.3 (c) Monthly average daily load curves; Riccarton Borough Council, April 1967 - March 1968

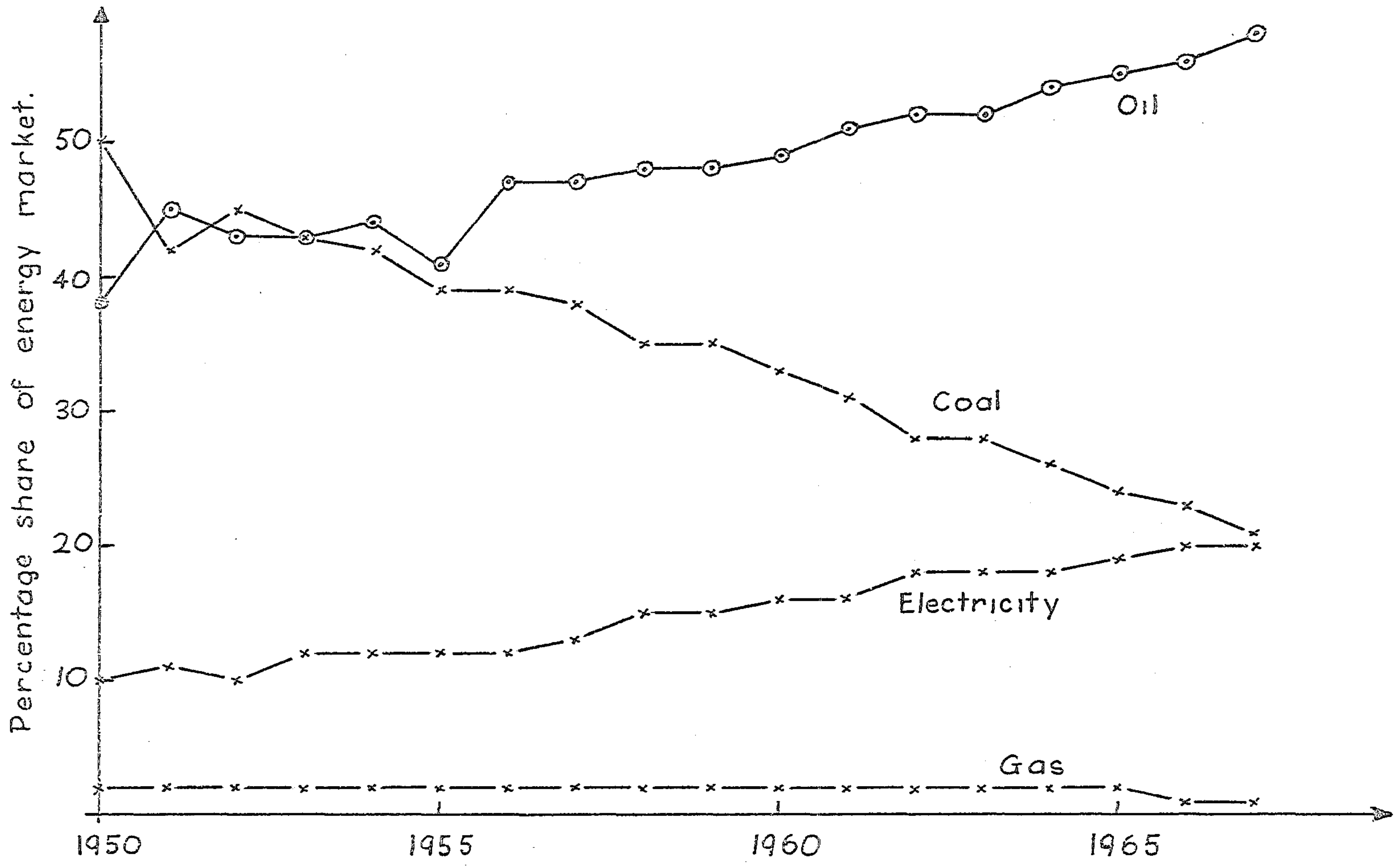


Figure 2.4 Share of New Zealand energy market met by 'major' energy sources.

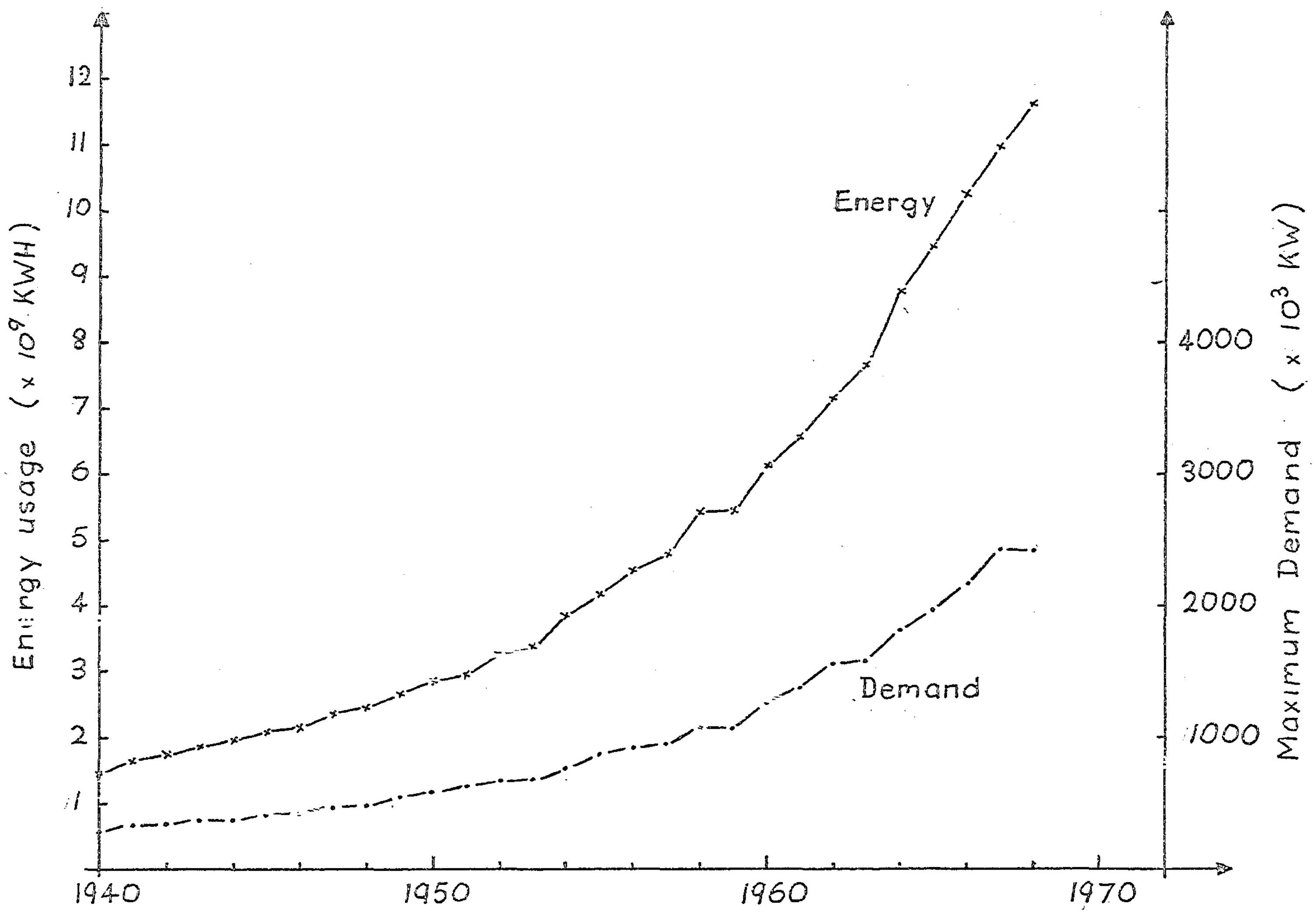


Figure 2.5 New Zealand annual energy usage and maximum demand.

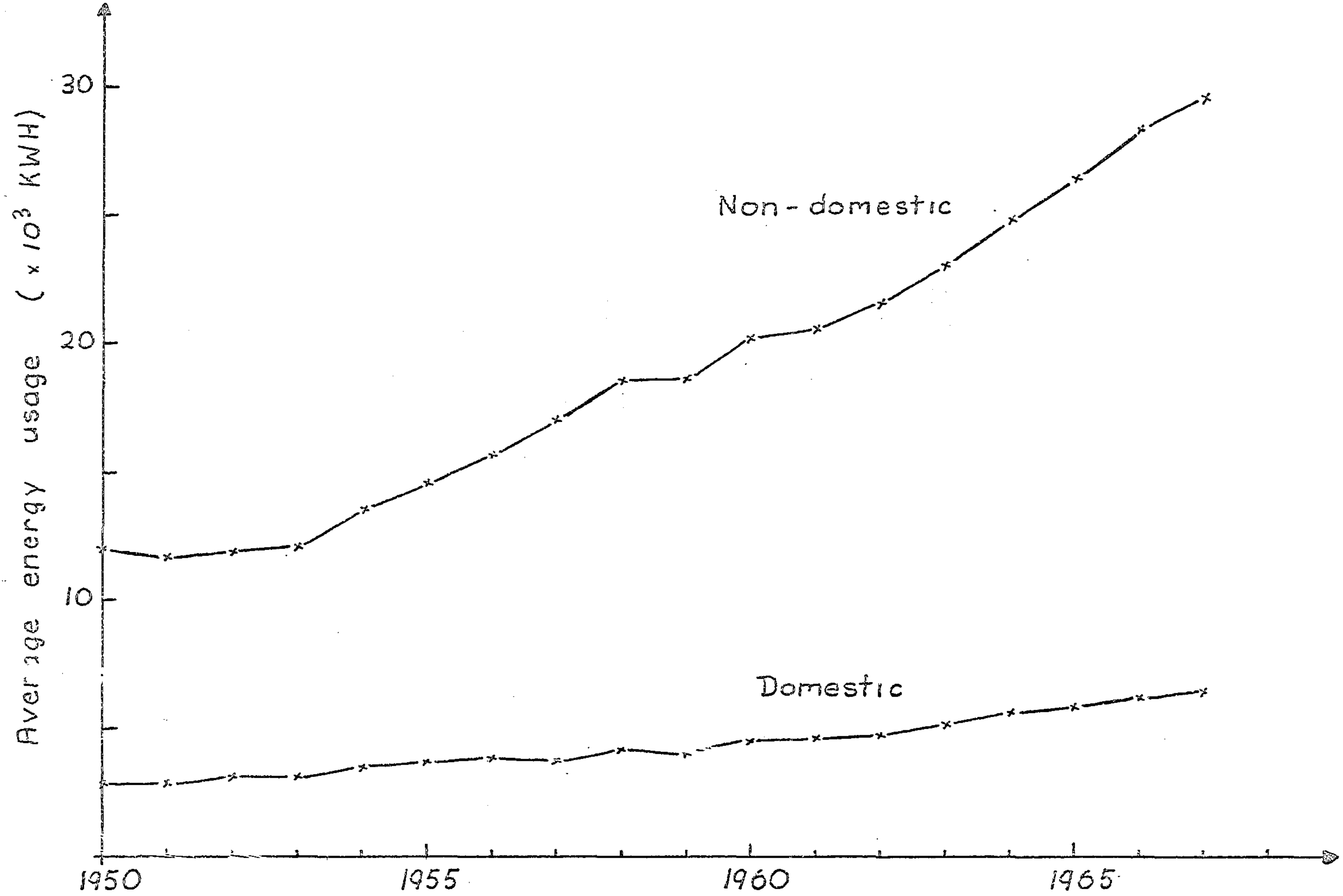


Figure 2.7 Average energy usage by domestic and non-domestic consumers in New Zealand.

CHAPTER 3

DESIRABLE MODELLING AND FORECASTING ACCURACY

3.1 Introduction

The accuracy required of a model of the behaviour of the demand is determined by the use to which the model is to be put; in this case forecasting future electric energy requirements. The supply of electric energy to the consumers must achieve, or exceed, a minimum level of reliability specified by management(16,35). Because equipment is not completely reliable and forecasts are not completely accurate some spare capacity or additional fuel storage must be provided to cover contingencies such as plant failure or forecast error (36). Spare capacity is, by nature, at least partly unproductive and represents a cost which management attempts to minimize while maintaining an acceptable level of reliability.

If demand, or energy, forecasts were more accurate less spare would be needed to allow for forecasting errors. Improvements to available forecasting methods have been directed toward reducing errors, or obtaining more comprehensive forecasts (1). The performance of a forecasting method is measured by the errors, i.e. actual value minus forecasted value, achieved over a number of experimental forecasts. The distribution of errors is then a measure of the "goodness" of the method. The method chosen from those available for a particular application is the one which performed with the least error (in some sense) over a number of experimental forecasts (1, 17).

Improvements to forecasting procedures cost money. No criterion appears to exist which determines whether a method is sufficiently accurate for the proposed application. In this chapter such a criterion is proposed. The basis for the criterion is whether the uncertainty in the forecast is small enough for the management target to be met. It is then shown that there is no advantage in improving demand forecasting accuracy, i.e. reducing the uncertainty, beyond limits set by plant unreliability and the level of reliability specified by management. Although the argument is developed primarily in terms of demand forecasts, i.e. of kilowatts, it may also be applied to energy forecasts as is pointed out in various places in the discussion.

3.2 The Criterion of Sufficient Accuracy

3.2.1 A measure of forecast uncertainty.

The error in a demand (energy) forecast can only be determined after the true demand (energy) value is known (36, 37). It is unlikely that future demands can be forecast with zero error in all cases. Consequently at the time when the forecast is made there is some uncertainty about what the true value will be. The future demand may be written

$$D = D_f + v \qquad (3.1)$$

where D_f = expected (forecast) value of demand in future

v = uncertainty.

The uncertainty is assumed to be distributed in some way with zero mean and variance σ_v^2.

To ensure that the risk of not meeting the demand (36, 38) is acceptably small it is necessary to provide additional generating capacity to allow for this uncertainty. Therefore a demand forecast should be accompanied by an estimate of the uncertainty such as upper, and lower, limits at some stated level of confidence (39). Ideally the form and parameters of the distribution of the probable demand is forecast.

Let v_z be the magnitude of the uncertainty at the z % confidence level and let G_T be the total capacity (including an allowance for uncertainty $G_v \geqslant 0$) provided to meet the demand; i.e.

$$G_T = D_f + G_v \tag{3.2}$$

and let $G_v = v_z$

When the plant is completely reliable the probability (or RISK) of not meeting the demand in full is

$$\text{RISK} = p\ (D > G_T) = p\ (D > D_f + v_z) \tag{3.3}$$

$$= \tfrac{1}{2}\ (1 - \tfrac{z}{100}) \text{ for a symmetrical distribution.}$$

The confidence level z is chosen to make RISK acceptable to management.

3.2.2 A Minimum Cost Criterion of Sufficient Accuracy.

Management also requires that the system be planned, or operated, in the most economical way. Reserve capacity is, by nature, unproductive and represents a cost which must be borne, and later recovered, in some way. This cost arises from fuel "wasted" in providing spinning reserve or from idle capital investment. Provided that the least costly method of obtaining a given reserve capacity is used this cost will not increase as the amount required decreases (curve (a) in figure 3.1). Furthermore it tends to zero as the reserve capacity tends to zero.

To minimize the reserve requirements the uncertainty, at the specified confidence level, must be reduced. This requires additional forecasting data, or additional processing of that already available, or both; hence an increase in costs. A reduction in uncertainty will, therefore, be accompanied by an increase in the cost of the forecast. If a completely different forecasting method is required the cost may rise abruptly. Uncertainty about future demand levels can be eliminated almost completely if all loads are controlled, which is clearly very expensive. Provided that the least costly method of forecasting to a given amount of uncertainty is used the cost of forecasting will increase as the uncertainty decreases (curve (b) of figure 3.1).

The total cost of the uncertainty is the sum of the costs of forecasting and of providing the necessary allowance (curve (c) of figure 3.1). There is no advantage in decreasing the uncertainty if this sum ceases to decrease. The criterion of sufficient accuracy is then:

> "A forecast is sufficiently accurate when a further reduction in the uncertainty, at the specified confidence level, is not accompanied by a reduction in the total cost of that uncertainty."

When plant is completely reliable this criterion specifies the amount of reserve required if the supply is to achieve the desired level of reliability in the most economical way. From figure 3.1 the total cost curve is a minimum at $v_z = v'_z$, giving $G_v = v'_z$. To achieve a specified RISK of, say, 0.001 requires, from 3.3, that $z \geqslant 99.8\%$. The uncertainty may reasonably be assumed normally distributed (14, 38, 40). Hence from tables of the standard normal distribution it is seen that

$$v'_{99.8} \approx 3.1 \ \sigma_v$$

Note that with G_v held constant an increase in z (reduction in RISK) requires a reduction in σ_v.

3.3 Reserve Capacity for Plant Outage and Forecast Uncertainty.

3.3.1 Plant outage reserves.

Plant is not completely reliable. Some reserve capacity is therefore provided to reduce the risk of not meeting the demand due to plant outage to an acceptable value. In fact, for many years plant outage was the only factor considered in reliability calculations (36, 38). Assume the forecast demand is certain to occur, i.e. $D = D_f$, $v_z \equiv 0$ for all z, and a set S of generators is scheduled to meet it. In this context "scheduled" refers to

daily generation scheduling or scheduling of installation of new generators, as appropriate. The total capacity of this set is

$$G_T = \sum_{i \in S} \hat{g}_i \tag{3.4}$$

where $\hat{g}_i$ = maximum capacity of generator i.

A reserve capacity G_o ($\geqslant 0$) is provided for in the schedule such that

$$G_T = D_f + G_o$$

Let s_{qj} denote the qth subset of the set S, such that the total generation available from the members of s_{qj} is G_j, where

$$G_j = \sum_{i \in s_{qj}} \hat{g}_i \quad , q = 1, 2, \ldots, Q_j \tag{3.6}$$

Q_j an integer $\geqslant$ 1. which denotes the number of subsets s_{qj} having a capacity G_j. Also, $j = 1, \ldots, n'$, where n' denotes the number of different values of G_j possible.

The G_j are ordered such that

$$G_1 = G_T > G_2 > G_3 > \ldots G_{n'-1} > G_{n'} = 0 \qquad (3.7)$$

A failure to meet the forecast demand in full occurs when any actually available set of generators, s_{qj}, has a generation capacity $G_j < D_f$. Using the Loss-of-Load Probability Technique, (12),

$$RISK_1 = \sum_{j=k}^{n'} \left(\sum_{q=1}^{Q_j} \prod_{i \notin s_{qj}} p_i \prod_{i \in s_{qj}} (1 - p_i) \right) \qquad (3.8)$$

where p_i = probability that generator i is not available, and

k = a number such that $G_j \geqslant D_f$ for $j = 1, \ldots, k-1$ and

$G_j < D_f$ for $j = k, \ldots, n'$

Sufficient reserve, G_o, must be available to maintain $RISK_1$ at an acceptable value while all members of the set S' are in operation. (Some degradation of $RISK_1$ will occur after the forced outage of a member). This is the minimum amount of reserve capacity that can be provided, regardless of the forecasting accuracy, if the level of reliability is to be acceptable. Furthermore the cost associated with G_o forms an irreducible lower bound on the cost of reserve capacity (curve (a) of figure 3.2).

3.3.2 Combined plant outage and uncertainty reserves.

This plant outage reserve, G_o, is required as such for only a portion of the time (e.g. if $RISK_1 \leqslant 0.001$ the full amount is required no more than 0.1% of the time). When it is not all required to cover plant outage it can be used as an allowance for uncertainty. It will not be necessary to provide additional reserve to cover the uncertainty if the whole of the required allowance can be obtained from G_o without increasing the risk to an unacceptable value. Hence the uncertainty at the z% confidence level must satisfy.

$$v_z \leqslant G_o \tag{3.9}$$

The expression for the probability of not meeting the demand in full now becomes

$$RISK_2 = \sum_{j=1}^{n'} p(D > G_j) \left(\sum_{q=1}^{Q_j} \prod_{i \notin s_{qj}} p_i \prod_{i \in s_{qj}} (1-p_i) \right) \tag{3.10}$$

For normally distributed uncertainty with

$$v_z = K_z \cdot \sigma_v$$

Where K_z = standard normal variate at the z% confidence level,

$$p(D > G_j) = \int_{G_j}^{\infty} \frac{1}{\sigma_v \sqrt{2\pi}} \exp\left(- \frac{(L - D_f)^2}{2 \cdot \sigma_v^2} \right) dL \tag{3.11}$$

Changes in the risk due to using the plant outage reserve additionally as an uncertainty allowance are thus a function of σ_v and, hence, the confidence level z. (Note that as $\sigma_v \rightarrow 0$, $RISK_2 \rightarrow RISK_1$). From equations 3.10 and 3.11 an upper bound on σ_v, and a corresponding lower bound on z, can be established if $RISK_2$ is to be acceptable for a given set S of generators and a given plant outage reserve G_o.

3.3.3 A minimum reserve criterion of sufficient accuracy

When no separate allowance for uncertainty is provided its cost is zero and the cost of the total amount of reserve provided remains constant as v_z increases (curve (a) of figure 3.2). However $RISK_2$ increases with v_z with a constant amount of reserve. At some value of v_z ($= v_z''$) $RISK_2$ reaches an acceptable value; thereafter additional reserve is required. From this point the cost of the total reserve, i.e. the sum of the plant outage reserve and this additional reserve, increases with v_z (curve (b) of figure 3.2). Provided the cost of forecasting (curve (c) of figure 3.2) is small at $v_z = v_z''$) the sum of these costs is a minimum at v_z''. There is then no advantage in reducing v_z below an amount equivalent to the minimum plant outage reserve G_o.

If the forecasting costs are not relatively small the total cost curve reaches a minimum at some $v_z > v_z''$. In this case the reserve capacity provided is equivalent to v_z and there is no advantage in reducing the uncertainty below this larger amount.

3.4 Criteria for System Reliability

From equation 3.10 and using the criteria of sufficient accuracy the accuracy with which it is desirable to forecast may be determined numerically, provided an acceptable level of risk is specified. At present there is no general rule available for specifying what level of risk is acceptable to management (40, 41, 42). In practice acceptable levels have been found by inserting the reserve capacities found satisfactory from experience into expressions such as equation 3.10 (16, 38, 40, 41). This has resulted in, for short term applications, probable shortage of capacity occurrences of from 0.001 to 0.0001 per hour (40), while for long term applications a probable shortage of capacity on 0.2 to 0.02 days per year appears satisfactory (41). These figures do not say a shortage will occur, merely that it is probable that so many will be experienced in a given period of time.

By itself a RISK measure gives no indication of the severity of a failure to meet the demand. It is possible that frequent outages, small in terms of the amount of capacity short, can be handled by operating the system at reduced voltage and frequency. Consequently attempts have been made to find more meaningful measures of unreliability, such as in the extensive work by Ringlee et. al. (43, 44) on the frequency and duration of outages. The expected magnitude of an outage may also be used as this gives a direct measure of the magnitude of the corrective action, e.g. load shedding, required if an outage occurs. Denoting the magnitude of an outage by u and the probability of its occurrence by p(u), the expected magnitude of an outage, when it occurs, is given by

$$E\{u\} = \int_0^{\infty} u\,\mathbb{p}(u)\,du \qquad (3.12)$$

where $$\int_0^{\infty} p(u) \; du = 1$$

A derivation of the expected magnitude of an outage based on equations 3.10 and 3.12 is given in Appendix A.

These measures do not solve the problem of deciding how much reserve capacity is needed in a particular situation; they merely form a basis for comparing differing amounts of reserve. Whether a particular amount of reserve gives satisfactory (to the consumers) reliability must be determined, at present, by experiment. If consumers complain then management must specify a higher level of reliability, which means more reserve and increased costs. This process may continue for ever; as consumers become accustomed to a higher level of reliability they will complain about smaller and smaller disruptions to the supply (16). Theoretically, it is possible to equate the cost of reserve to the cost of a disruption of the supply and hence find a minimum cost solution to the problem. Practically, there are extreme difficulties in adequately determining the cost of a failure to meet the demand, except in a few very special cases (45).

Consequently the traditional criteria for determining the amount of reserve required must still be used. These give reserve margins:

(a) as a fixed percentage of the load, or

(b) equivalent to the capacity of the largest generating unit.

Combinations of both may be used together with a probability analysis (36, 46, 47).

In the following sections the numerical evaluation of sufficient accuracy based on traditional reserve criteria is discussed for short and long term demand forecasts.

3.5 Accuracy of Short Term Demand Forecast

3.5.1 Purpose of short term demand forecasts and minimum lead time.

The demand varies continuously throughout the day. Generation capacity to meet it is effectively variable only in discrete steps. There is a time lag, which varies with the type of plant, between a generator being started and its coming "on load". On the basis of the short term demand forecast generation is scheduled for starting so that sufficient capacity, including spinning reserve, is available when required. This spinning reserve is generally allocated to a number of machines either strategically or in accordance with costs (16, 46). Any generation schedule may be updated on the basis of revised demand forecasts provided there is still time to start up extra generators. Short term forecasts must, therefore, achieve the desirable accuracy over lead times not less than this starting time.

Short term random excursions of the demand about that forecast are met by the spinning reserve. However, persistent trends away from the forecast, e.g. as a result of changed weather conditions, This page intentionally blank.

can only be met by starting up extra plant. The warning required for such action varies from a few minutes for hydro or gas turbine plant up to 12 hours for thermal plant. Unscheduled starting of additional thermal plant in particular is to be avoided if possible because of high starting costs.

3.5.2 Spinning reserve requirements

With insufficient spinning reserve undesirable system behaviour, e.g. instability, unacceptable frequency drops etc., may follow the sudden loss of generating capacity which accompanies a forced outage. Load shedding may be necessary if such behaviour is to be corrected, or complete system collapse may follow. For small magnitude outages it may be acceptable to operate the system at reduced voltage and frequency.

Outage conditions, i.e. $G_j < G_T$, in the short term case, exist only until the failed generator(s) can be replaced or repaired. (Replacement is normally faster than repair). The probability p_i, of section 3.3.1, that generator i, $i \in S$, is forced out of service is interpreted as the "unreplaced outage rate"(35, 40); i.e.

$$p_i = x \, . \, \tau \qquad (3.13)$$

where x = frequency of forced outages (occurrences per hour), and

τ = mean time to replacement (hours)

Typically for hydro x is about 0.0001 - 0.0005 failures per hour, independent of capacity (48). Replacement times are much less than one hour giving p_i about 0.0001 or less. For thermal plant x varies with plant size from about 0.0005 up to 0.002 for large new plant (49). With replacement times of two or more hours p_i can vary from about 0.001 to 0.01 in the case of very large new generators.

In New Zealand the management criterion for spinning reserve is that sufficient be provided to avoid shedding load in at least the more probable outage conditions. This is similar to many other cases (16, 50). In practice "more probable" has usually been interpreted as the probability of forced outage of any one generating unit. The probability of any one unit failing is of order p_i, while the probability of two units failing simultaneously is of order p_i^2. Hence one unit failing is "more probable" than the simultaneous failure of two units by a factor $(^1/p_i)$. For hydro plant this factor is of order 10,000, dropping to about 1000 for thermal plant.

3.5.3 Evaluation of desirable forecasting accuracy.

Clearly it is difficult to evaluate the risk of not meeting the demand, and hence compare reserve schemes, without reference to a particular system. For the purposes of illustration consider the simple hydro system of figure 3.3 and table 3.1; this is based on that of the South Island of New Zealand. Plotted in figure 3.4 are the results of calculations of the risk of not meeting in full forecast demands varying from 1080 MW to 960 MW when all generators are scheduled to operate, i.e. G_T = 1080 MW and G_o varies from 0 to 120 MW. Curve (a) shows the variation in risk assuming the forecast demands are certain to occur, i.e. v_z = 0 for all z.

The curves (b), (c) and (d) are evluations of $RISK_2$ (equation 3.10) for three values of z, the confidence level on the distribution of v, and with

$$\begin{aligned} v_z &= G_o \\ &= G_T - D_f \\ &= K_z \cdot \sigma_v \end{aligned} \qquad (3.14)$$

The three curves are plotted for values of $K_z = 3, 4, 5$. Assuming the uncertainty is normally distributed these correspond to levels of confidence of 99.75%, 99.992% and 99.9998% respectively.

If the specified plant outage reserve (see section 3.5.2) is given by

$$G_o = \hat{g}_i = \max (g_i) \;,\; i \in S \qquad (3.15)$$

then the acceptable risk of not meeting the demand must, from figure 3.4, be less than 6×10^{-4}, otherwise less spinning reserve could have been used. Now if G_o is to be used additionally to cover forecast uncertainty $RISK_2$ (equation 3.10) must also be less than 6×10^{-4}. To achieve this K_z must have a value of about 4, or greater; see figure 3.4.

Hence in the absence of an analytic method of determining a satisfactory level of risk the following statement can be made about the accuracy of short term demand forecasts;

"A short term demand forecast is sufficiently accurate if the uncertainty, at a confidence level of about 99.99% or greater, is equivalent to the capacity of the largest generator in operation."

3.6 Accuracy of Long Term Demand Forecasts

3.6.1 The purpose of long term demand forecasts.

The installation of extra generating capacity to meet the annual increase in maximum demand is planned, i.e. is scheduled, on the basis of the long term demand forecast. The installed capacity should be sufficient to meet the expected maximum demand and provide any necessary spinning reserve. The generating capacity margin set at the planning stage fixes the available capacity, and hence the operating margin at the time of peak demand, for several years into the future.

3.6.2 Installed capacity reserve margins.

When estimating the long term risk of not being able to meet the demand in full the probability p_i is interpreted as the probability that generator i is not available for use, as distinct from the probability that it will be forced out of service. A generator is not available for use when it is being maintained either on a planned basis or on an emergency basis. Planned maintenance is normally done during periods of light load, i.e. during the summer in New Zealand, and consequently at the time of maximum demand all generators are ideally available for service. Consequently p_i may be taken as the probability that generator i is on forced outage at the time it is required. The probability that the demand will not be met is thus a measure of the risk that there will be insufficient installed capacity at the time of maximum demand (41).

Typically hydro plant is on forced outage 0.25% to 0.5% of the time (48). This figure is higher for thermal plant, being up to 10% for large new plant (49, 52).

As the largest generators in a power system are usually also the newest they are, in general, more likely to be on forced outage at the time of maximum demand. Furthermore, at this time, all generators will be scheduled for operation and hence it is not possible to start up more plant. Consequently the installed reserve capacity is, as a minimum, equivalent to the capacity of the largest installed generator (16, 46, 52). In New Zealand the aim is to maintain an installed reserve equivalent to the capacity of the two largest thermal machines, but this is not always possible (51).

3.6.3 Desirable long term forecasting accuracy.

When the installed reserve capacity is equivalent to the capacity of the largest generator the probability of not being able to meet the full demand is again of order $(p_i)^2$. A probability that the demand will not be met in full, about once in twenty years, appears to be a commonly accepted level of reliability (41, 49). The additional use of the installed reserve to cover forecast uncertainty should not increase the risk above this figure. Hence the confidence level on the forecast, z, should be about 95%, assuming normally distributed uncertainty.

The following statement can therefore be made about the accuracy of long term demand forecasts:

> "There is no additional gain from reducing the uncertainty at a confidence level of about 95%, below an amount equivalent to the capacity of the largest generating unit in the system."

This accuracy must also be achieved over a minimum lead time equivalent to the time required to install new generating plant, normally two or more years.

3.7 Energy Usage Forecasts

3.7.1 A criterion of desirable accuracy of energy usage forecasts.

Energy usage forecasts determine, at the planning stage, the amount of plant and water storage required in hydro systems or the plant mixture in hydro/thermal systems. The seasonal nature of the water inflow pattern dictates that an "energy balance" be obtained over one seasonal cycle (occasionally two for large reservoirs). For convenience the length of this cycle is taken as one year (365 days). This is a fixed, and known, length of time for planning purposes. This "energy balance" is achieved by storing water in sub-periods, e.g. months, when inflow is high and energy usage low (summer) and releasing extra water in the winter when the opposite occurs.

This "energy balance" is illustrated in figure 3.5. The predicted energy requirements in each sub-period (curve (a)) are met from the available inflow (curve (b)) and the accumulated stored water (curve (c)). Any error in the energy usage forecast will result in the accumulated surplus being used at a greater rate, if the forecast is low, or will increase the accumulated surplus if the forecast is high. An error in the annual energy usage forecast must be met from the net surplus of water over the year. This is obtained from excess inflow, plus any accumulated stored water during the year plus any surplus from the previous year. Additional water storage is provided in hydro systems so that surplus water from "normal flow" years can be stored and used in years when conditions are such that probable minimum flows occur, to ensure continuity of supply.

A criterion of desirable accuracy for annual energy usage forecasts is then:

> "There is no additional gain from reducing the uncertainty in the annual energy usage forecast, at a specified confidence level, below that amount of energy obtainable from the net surplus of water, integrated over the year, available when water inflows reach their probable minimum."

In this context "probable minimum" means the lower confidence limit on the distribution of water inflows.

3.7.2 Available energy forecasts and the accuracy of energy usage forecasts.

In any particular period of time an estimate of the amount of energy which will be available is obtained from the forecast inflow patterns, in conjunction with the available storage. There is uncertainty about how much energy will be available. It is of interest to determine the effect of this uncertainty on the desirable accuracy of energy usage forecasts. The energy usage forecast is assumed to have a probability density function $P_r(W)$; $P_a(W)$ is the available energy density function. The probability that the available energy W_a is less than some value W_1 is

$$P(W_a < W_1) = \int_{-\infty}^{W_1} P_a(W)\, dW$$

The probability that the required energy lies between W_1 and $W_1 + dW$ is

$$P(W_1 < W_r < W_1 + dW) = P_r\,(W_1)\, dW$$

Hence the probability that the required energy exceeds that available is

$$P(W_r > W_a) = \int_{-\infty}^{\infty} P_r(W_1) \left[\int_{-\infty}^{W_1} P_a(W)\, dW \right] dW_1 \qquad (3.16)$$

This result holds for energy forecasts over any time interval.

The two distributions may be assumed normal in many situations, i.e. $P_r(W) = n(\mu_r, \sigma_r^2)$; $P_a(W) = n(\mu_a, \sigma_a^2)$

Hence equation (3.16) may be written as

$$P(W_r > W_a) = \int_{-\infty}^{\infty} \frac{1}{\sigma_r\sqrt{2\pi}} \exp\left(-(W_1 - \mu_r)^2 / 2\sigma_r^2\right) .$$

$$\cdot \left[\int_{-\infty}^{W_1} \frac{1}{\sigma_a\sqrt{2\pi}} \cdot \exp\left(-(W - \mu_a)^2 / 2\sigma_a^2\right) dW \right] dW_1 \qquad (3.17)$$

On substituting $x = (W_1 - \mu_r) / \sigma_r$ equation (3.17) becomes

$$P(W_r > W_a) = \tfrac{1}{2} + \tfrac{1}{2} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \operatorname{erf}(\gamma_x + \beta)\, d\,x \qquad (3.18)$$

where $\gamma = \sigma_r / \sigma_a\sqrt{2}$, $\beta = (\mu_r - \mu_a) / \sigma_a\sqrt{2}$

In figure 3.6 the probability that energy requirements exceed that available is plotted against γ for several values of β. Note that σ_r and σ_a are measures of the deviations from the respective means and thus a measure of the uncertainty in the associated forecasts. When the mean values occur the quantity $(\mu_r - \mu_a)$ represents the amount of excess energy required (if $\mu_r > \mu_a$) or which may be stored (if $\mu_r < \mu_a$).

Figure 3.6 indicates that for a given $P(W_r > W_a)$ as γ decreases then $|\beta|$ can be decreased. This means the available water can be utilized more efficiently. The curves also show there is no lower limit on γ and hence on σ_r but there is a lower limit on $|\beta|$. For a 10% risk of failure to supply the full amount of energy required this limit is a little less than 2. To achieve this risk level with $|\beta| = 2$ then $\sigma_r < 0.707\ \sigma_a$. There is a maximum limit on the uncertainty in the energy forecast, for a given separation of the distributions, but no lower limit in the same sense as in the peak demand case.

3.8 Summary

A criterion of sufficient accuracy for forecasts of demand and energy usage has been described. Sufficient accuracy has been defined in terms of the cost of the spare capacity required to ensure that the risk of not meeting the demand, or of being unable to meet energy requirements, is less than some value specified by management.

It has been shown, using this criterion, that there is no advantage in reducing the uncertainty in the demand forecasts, at a specified confidence level, below an amount equivalent to either the largest generator in operation (for short term forecasts) or the largest generator installed (for long term forecasts). A simple example was used to illustrate the argument in the short term demand case.

A separate criterion was obtained for the desirable accuracy of energy usage forecasts. This was in terms of the amount of surplus water available. The effect of uncertainty in forecasts of the amount of water available on the desirable accuracy of energy usage forecasts was discussed.

These criteria imply that the amount of uncertainty in a forecast is known, or can be estimated. In the next chapter several common forecasting methods are discussed from the point of view of determining the uncertainty.

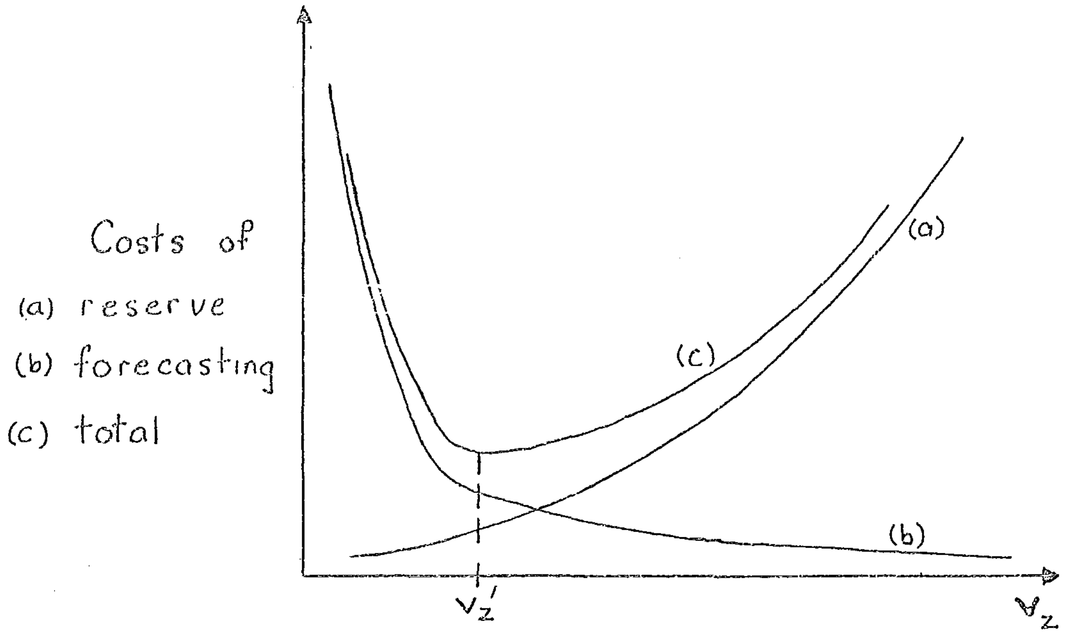


Figure 3.1 The Cost of Uncertainty

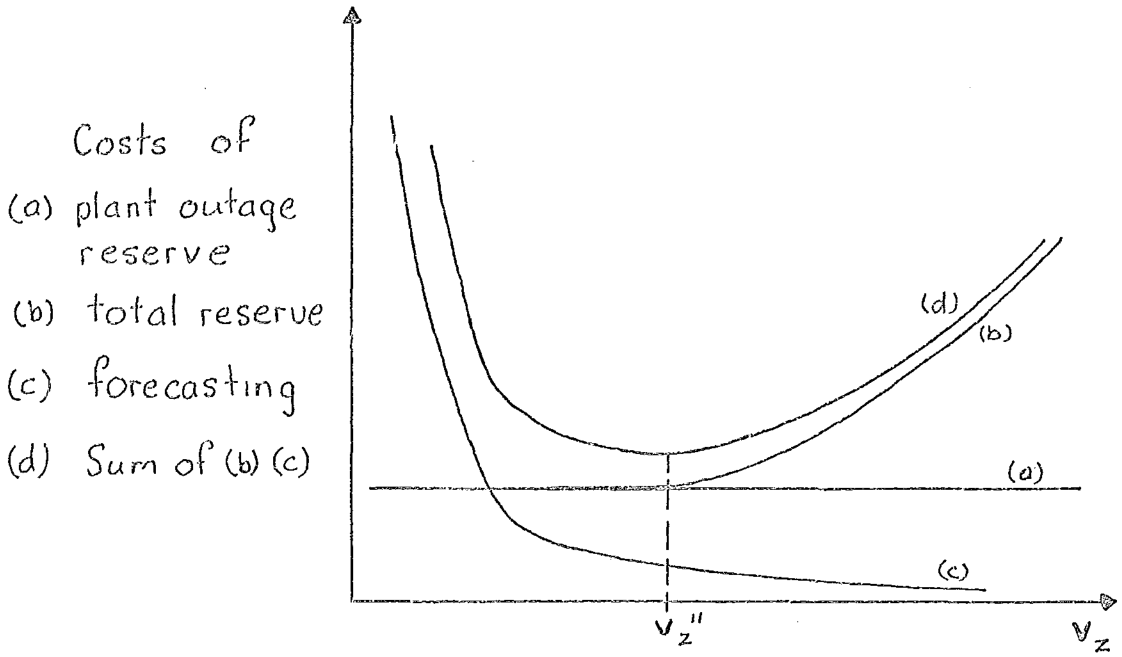


Figure 3.2 The Costs of Uncertainty and Plant Outage Reserve

TABLE 3.1

GENERATOR PARAMETERS

UNITS	CAPACITY (MW)	FREQUENCY OF FORCED OUTAGE (OCCURRENCES/HOUR)
1 - 6	90	0.0001
7 - 10	55	0.0001
11 - 18	40	0.0001

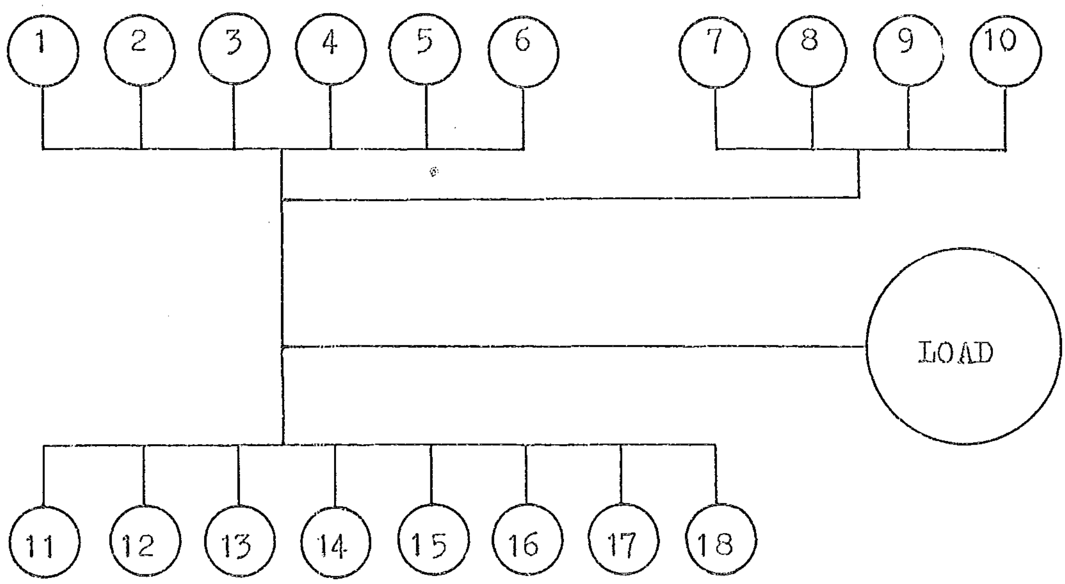


Figure 3.3 Simple 18 generator and single load power system.

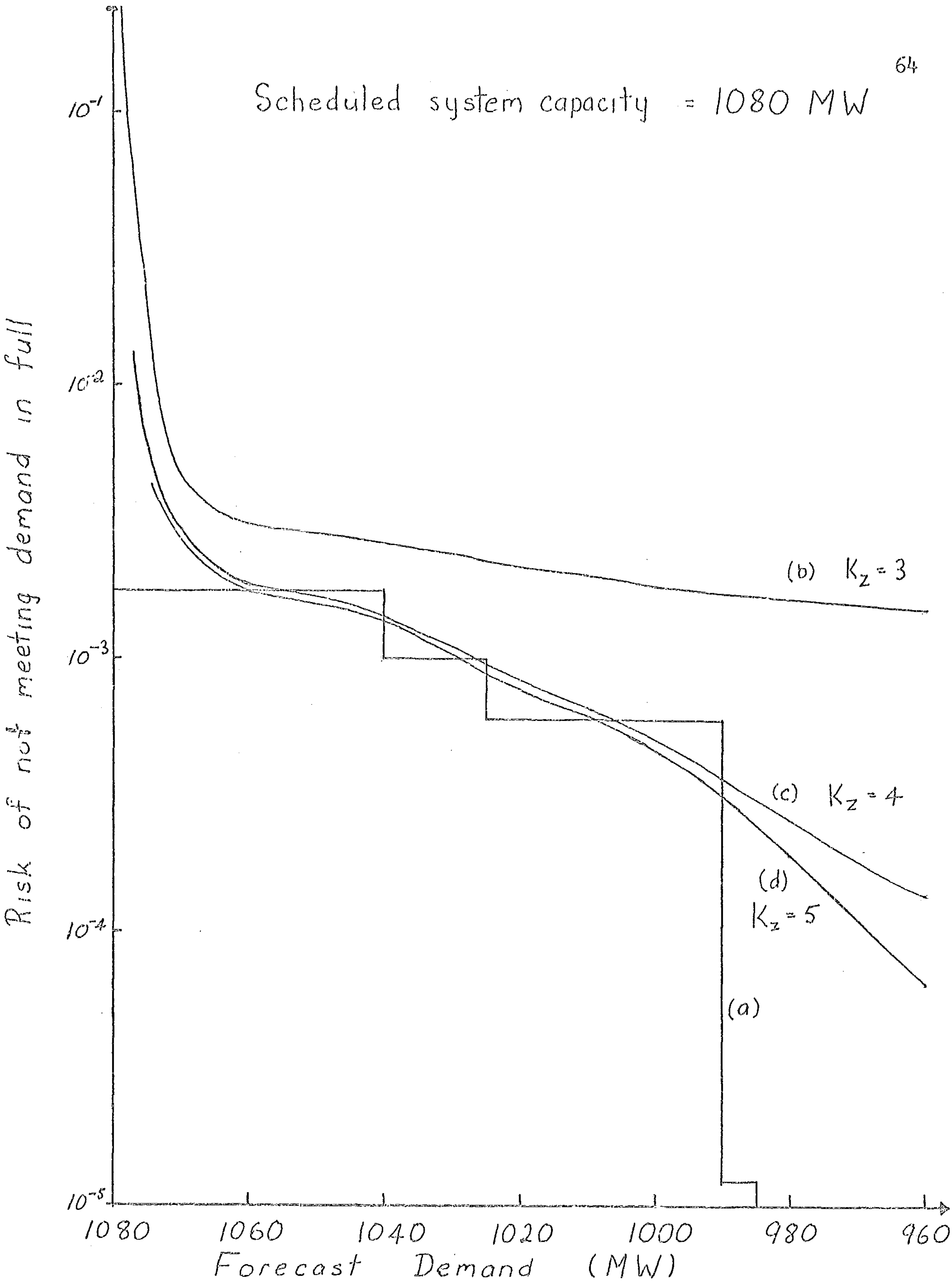


Figure 3.4 Risk of not meeting forecast demands.

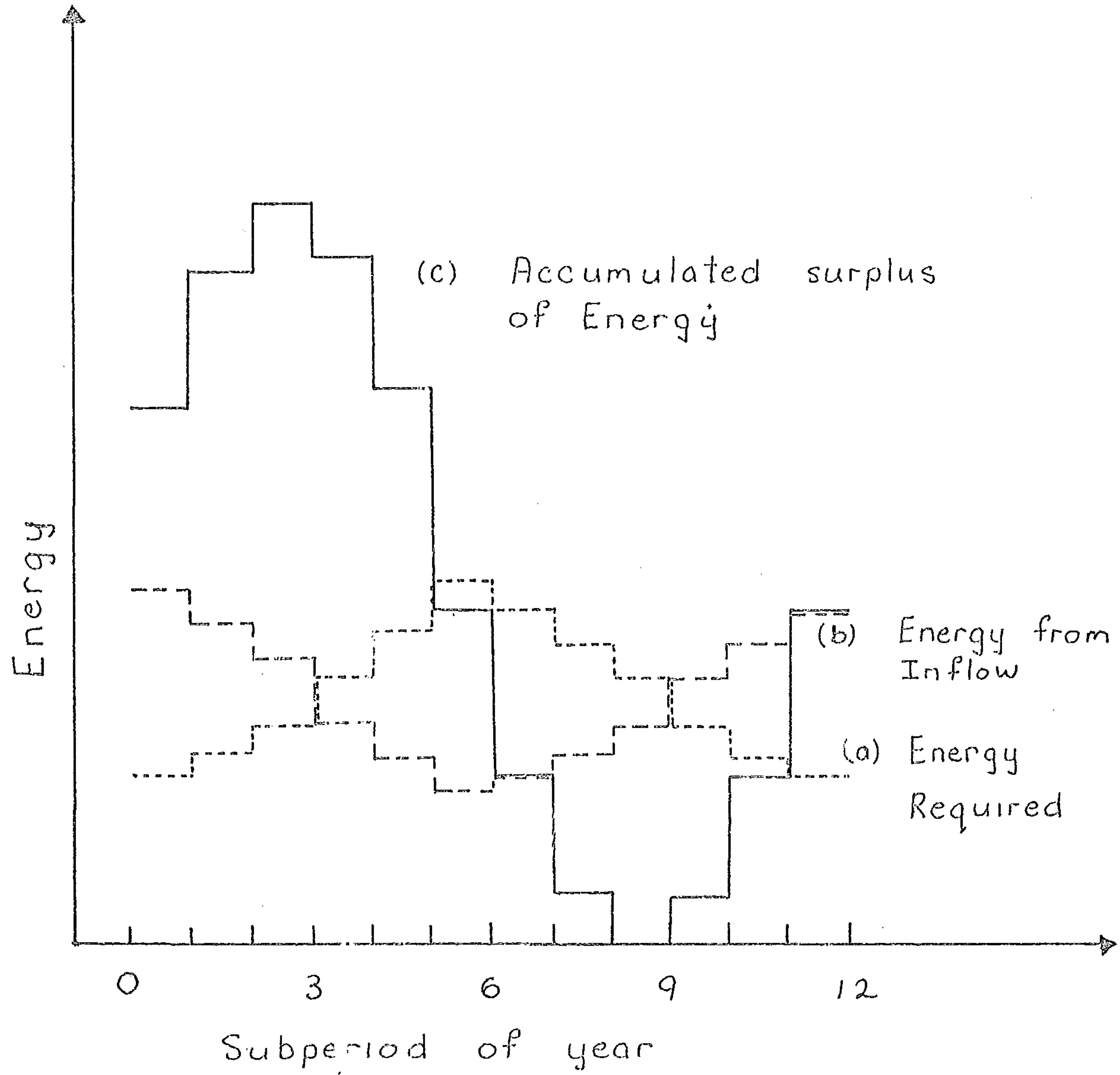


Figure 3.5 Seasonal Energy Balance

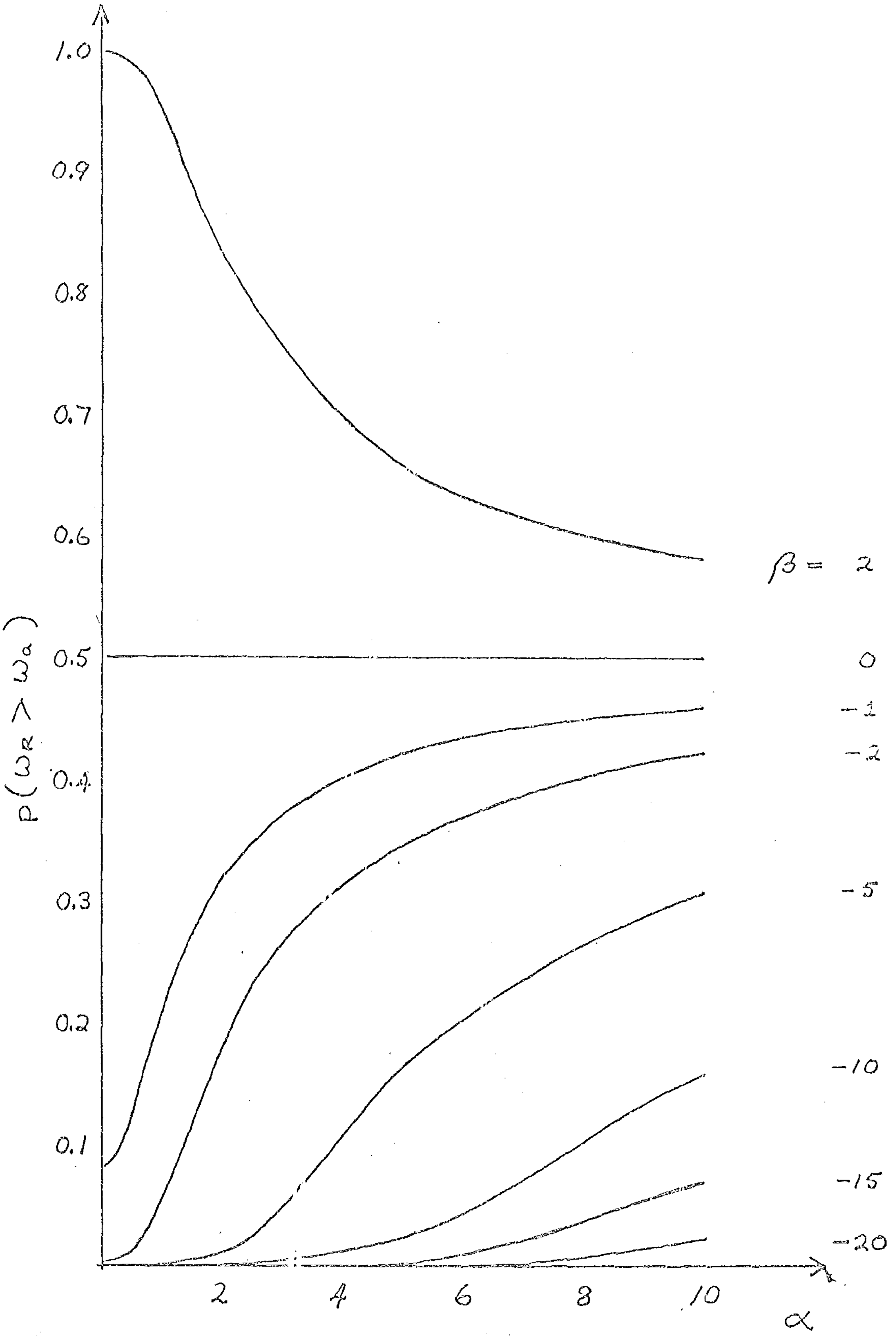


Figure 3.6 Probability of Required energy (ω_R) exceeding Available energy (ω_a)

CHAPTER 4

THE STATE OF THE ART

4.1 Introduction

Many methods have been developed for forecasting the demand for electric energy (1,2). Most have been tailored to a particular power system. This chapter critically reviews the major methods in use at present. Some methods which have been proposed but not used extensively are mentioned.

As the demand for electricity is a non-deterministic process no method can hope to forecast future demands exactly. Therefore this chapter pays particular attention to the estimation of uncertainty in forecasts made using the methods discussed. The viewpoint taken here is that a forecast for which estimates of the uncertainty are given is to be preferred over one for which these estimates are not available (4).

Examples of the major forecasting methods are discussed under three headings;

(a) Long term forecasting methods,

(b) Weather-load models, and

(c) Short term forecasting for system operation.

Long term forecasts, i.e. those over lead times of one or more years, are used primarily for system planning. Over these lead times the most significant feature of the load is its growth. Thus models of the growth process form the basis of all long term forecasting methods, whether for forecasting annual maximum demand or annual energy usage.

Weather-load models have been used for short term, i.e. lead times up to one year, demand forecasting (2, 14, 30) and for "explaining" fluctuations in the historical record of annual maximum demand (57, 59). Consumers vary their demands to minimize personal discomfort due to changes in the local weather. These models describe the variations in demand as functions of the weather variables relevant to the particular region. At present it appears difficult to achieve the sufficiently accurate weather forecasts which are necessary for accurate demand forecasts.

Much of the data required by weather load models, i.e. weather records and forecasts, is not readily available to the system operator. Consequently several short term forecasting methods which require only readily available demand data have been developed (10, 37, 72). These extrapolate trends in the daily load curve. It is suggested that these methods would be improved by the inclusion of known weather information (which may be obtained relatively easily by the system operator).

4.2 Long term forecasting methods

4.2.1 Trend extrapolation using least squares.

In the trend extrapolation methods an empirical model of the growth process is assumed (1, 11, 17, 38, 56, 59). This model represents the growth of demand as a function of time alone. From this model a mathematical trend curve (which includes unknown parameters) is obtained. The parameters are found by fitting the trend curve to historical data. Forecasts are obtained by extrapolating the fitted trend curve into the future. It is assumed that the same model is a valid approximation to both past and future behaviour of the growth process.

Several models of the growth process have been used, depending on the shape of the plot of demand against time (some examples of the resulting trend curves are given in table 4.1). Of these models the simplest (and most widely used) assumes the rate of growth proportional to the demand already in existence, i.e. that growth occurs exponentially. Using the notation of table 4.1 this model is written

$$\frac{dD}{dx} = b_o D$$

The trend curve is then written as

$$Y_t = a_o + b_o (x_t - \bar{x}) = (\text{Log}_e D_t) - e_t \qquad (4.1)$$

where $\bar{x}$ denotes the mean of the independent variable over the relevant historical period; i.e. for N observations.

$$\bar{x} = \frac{1}{N} \sum_{t=1}^{N} x_t$$

Estimates of the parameters a_o, b_o are obtained using the method of least squares (26, 60) to fit the trend curve to N historical data values (usually annual figures).

Let y_t denote an observation of Y_t in period t (i.e. $y_t = \log_e D_t$). Least squares requires that the observations, y_t, be assumed,

(1) stochastically independent,

(2) distributed in some way about a mean value, Y_t, given by equation 4.1 (or any of the alternative models in table 4.1), and

(3) to have constant variance for all x_t (the variance can be proportional to a known function of x_t with increased mathematical complexity).

Leaving aside discussion of the validity of these assumptions for the moment, estimates of a_o, b_o are given by

$$\hat{a}_o = \frac{1}{N} \sum_{t=1}^{N} y_t$$

$$\hat{b}_o = \sum_{t=1}^{N} y_t \, (x_t - \bar{x}) \Big/ \sum_{t=1}^{N} (x_t - \bar{x})^2 \qquad (4.2)$$

Estimates of Y_t (equation 4.1) are obtained from;

$$\hat{Y}_t = \hat{a}_o + \hat{b}_o \, (x_t - \bar{x}) \qquad t = 1, \ldots, N$$

A forecast of the demand at time $N + \tau$, i.e. for a lead time of τ, is obtained from

$$\hat{D}_{N+\tau} = \exp (\hat{Y}_{N+\tau})$$

$$= \exp (\hat{a}_o + \hat{b}_o \, (N + \tau - \bar{x})) \qquad (4.3)$$

Uncertainty as to the true value of Y_t exists because $\hat{a}_o$, $\hat{b}_o$ are estimates of the true values and because the y_t are distributed in some way about Y_t (26, 60, 95). Assuming additionally that this distribution is normal the variance of the distribution of y_t is written

$$\text{var } (y_t) = \hat{\sigma}^2 \left(1 + (1/N) + (x_t - \bar{x})^2 \Big/ \sum_{i=1}^{N} (x_i - \bar{x})^2 \right)$$

$$\hat{\sigma}^2 = \frac{1}{N-2} \sum_{i=1}^{N} (y_t - \hat{Y}_t)^2 \qquad (4.4)$$

The range of probable demand, at the desired confidence level, is found from equations 4.3 and 4.4; i.e.

$$\exp\left(\hat{Y}_t - k\sqrt{\mathrm{var}(y_t)}\right) \leq D_t \leq \exp\left(\hat{Y}_t + k\sqrt{\mathrm{var}(y_t)}\right) \qquad (4.5)$$

where k denotes the 't' statistic at the desired level of confidence.

From a forecasting viewpoint the second of the least squares assumptions above (i.e. the assumed behaviour with time of Y_t) is the most critical. When the assumed growth model is not a reasonable approximation to the actual behaviour of the growth process, i.e. Y_t is not the mean of the observed y_t, then it is unlikely that the other least squares assumptions are valid. Furthermore if the assumed model contains too many unknown parameters there may not be sufficient data available to enable reliable parameter estimates (26, 60 95).

The assumption of exponential growth, as in equation 4.1, is a reasonable one. There is evidence for it in

(a) the time behaviour of historical demands,

and

(b) a population's tendency to grow in an exponential way (and it is also reasonable to assume that the additional population will seek a similar, or better, standard of living than the existing population and thus demand growth will be at least proportional to population growth - see chapter 2).

As there are only two unknown parameters in this simple trend curve reasonably good estimates can be obtained from quite small amounts of data.

It is sometimes assumed that the demand for electricity will saturate (cease to grow) at some time in the future, e.g. (9, 59). The basis for this assumption is that demand "cannot possibly" continue to grow at the historical rate (about 7%) (38, 59). Consequently a Gompertz or Logistic trend curve is selected for the model, Y_t. Such models as these assure eventual saturation of the forecast demands, regardless of the physical situation. As there is no physical evidence of the demand tending toward saturation now, or over the lead time of present forecasts, the use of saturating trend curves cannot be justified. Where they have been used the estimated parameter values are such that over the historical period and for three to five years into the future these curves are closely approximated by a simple exponential anyway (59). Where saturation is suspected it is more valid to treat the rate of growth as a function of time. For example Hore (38) suggests a "modified exponential" growth rate, giving a trend curve of the form

$$Y_t = a_1 + b_1 (x_t - \bar{x}) + c_1 (x_t - \bar{x})^2 \qquad (4.6)$$

Again the coefficients a_1, b_1, c_1 are estimated from historical data using least squares. If the hypothesis $c_1 = o$ is tested, statistical evidence for, or against, saturation can be obtained. Acceptance of the hypothesis, at some stated confidence level, means there is no statistical evidence of saturation. However it is possible for the hypothesis to be rejected if, in the previous one or two years, the growth rate has temporarily decreased due, perhaps, to a mild winter or economic recession. This fact emphasizes the need for professional judgement to be exercised before any statistical results are accepted, see also (11, 17, 26, 38).

In time series which are based on economic phenomena, and the demand for electricity is one, there is often considerable correlation between successive observations (26, 96). Consequently the first least squares assumption (of stochastically independent observations) is not always valid, particularly if the growth of demand exhibits some cyclic behaviour. The effect of such serial correlation is to reduce the number of degrees of freedom available when estimating the model parameters (95, 96). Thus if serial correlation is present but is ignored the estimates of the variance of y_t about Y_t will not be reliable. In some cases these effects can be minimized by modelling the growth rate rather than the trend, as done by Henault et.al (50), i.e. in the model

$$D_t - D_{t-1} = (b + e_t)\, D_{t-1}$$

(in which e_t is stochastically independent and distributed in some way about a mean of zero with variance σ^2), the parameter b is estimated via least squares from historical growth rates. However, the improvement in the reliability of the variance estimates is marginal if the growth rate varies cyclically.

The technique of trend extrapolation assumes that the growth process which gives rise to the observed demands

(a) remained the same throughout the relevant historical period, and

(b) will remain the same over the specified lead time of the forecasts.

Industrialization, changing living standards etc., mean that the growth process remains the same for only limited periods of time. If the N historical values are taken, as is usual, at annual intervals then only 10, or occasionally 20, data values can reasonably be used. Professional judgement must be used in determining how much historical data may be used in the parameter estimation process. (A good example of a limitation on N occurs in the New Zealand situation. Prior to 1959 energy rationing (power cuts) were used to conserve water. As a result consumers could not rely entirely on electricity as an energy source. Since 1959 when rationing ceased consumers have come to rely entirely on electricity, thus changing the nature of the growth process. Historical data since 1959 reflects unrestricted demand while before that year the values of energy usage and maximum demand have been artificially reduced. Consequently only data from 1959 onward can reasonably be used for extrapolation into the 1970's).

It is unreasonable to assume that the growth process will remain the same for longer in the future than it has in the past. Consequently the maximum lead time for forecasts obtained by extrapolation is limited to about 10 years, or less. Extrapolation past this limit is based on conjecture supported by judgement. Such forecasts need to be checked for feasibility against the demand achievable with foreseen technology (or with more "developed" countries) (61).

The small amount of data available (10-20 annual observations of demand) is not sufficient to form highly reliable variance estimates and hence of the uncertainty in the forecasts (1, 95). Several demands in any one year are potentially the annual peak because similar physical conditions occur in those days. These demands may be used in simple trend models to effectively increase the amount of data available without an increase in the number of years spanned. Thus the reliability (in a statistical sense) of the parameter and variance estimates could be improved (1, 77).

The above technique cannot be used to increase the amount of energy usage data as there can only be one energy usage value in each time interval. However it is possible to extract the long term growth trend from monthly energy usage observations by the successive application of weighted moving averages (11, 90, 97). In this way 12 data values per year are obtained. Forecasts of demand are obtained by applying trend extrapolation techniques to the monthly seasonally adjusted data series (11, 97). Seasonal adjustment by the successive application of moving averages, while widely used for adjusting economic time series, must be treated with caution, particularly as it may introduce "artificial seasonality" which is not present in the original series (98). Also some growth information may be removed along with seasonal variation (63, 98).

Provided the model selected to represent the time behaviour of the demand data is appropriate the accuracy with which the future can be forecast is a function of the fluctuation of the data about the assumed model. Thus the adequacy of any particular trend extrapolation method (with, or without, seasonal adjustment etc) can only be determined through experimentation with actual data. The model of equation 4.6 was used to make forecasts of NZED annual maximum demand and annual energy usage; the parameter c_1 was included if the hypothesis $c_1 = 0$ was rejected at the 95% confidence level. The most recent ten annual values of demand were used to estimate the parameters. Forecasts for lead times of one to five years are shown in tables 4.2 and 4.3, together with the percent uncertainty at the 95% confidence level. Using the demand estimates (in table 4.2) the required capacity margin at the 95% confidence level is considerably greater (at 5 - 20% dependent on lead time) than the 3 - 4% that is equivalent to the capacity of the largest generator (see chapter 3). Consequently this particular model cannot be considered adequate with respect to the desirable accuracy.

In the trend extrapolation methods discussed in this section time is the only independent variable. Consequently none can be used to determine the effect of proposed load control measures such as tariff modification. A more serious weakness in these methods is that if a change in the nature of the demand occurs (e.g. a new and popular type of appliance comes on the market, such as the range of 2 kW radiators in 1962-63) several time periods must elapse before the forecasts reflect this change. In many cases it is known (from market observation) that the change is occurring and ad hoc procedures are devised to modify the extrapolation to allow for the change.

4.2.2 The "Market research" approach.

The composition of the load changes with time due to changing living standards, economic conditions etc. Changing composition and the effect of variables other than time on the energy usage are taken into account in a "market research" approach, i.e. the question

"How much energy will we be able to sell in the next year?" is asked (12, 17). The inclusion of factors other than time in the model is expected to reduce the amount of unexplained variation of demand.

Hooke (17) forecast domestic sales from extrapolations (both mathematical and based on experience) of sales per consumer multiplied by predicted consumer numbers obtained from trends in the ratio of population to dwellings. Industrial sales, which in this case formed about 60% of total energy sales, were related to the long term trend in industrial development, based on selected indices of production. Alternatively sales were related to the size of the work force in the region and forecast from estimates of the future work force (see also Baker (31)).

The seasonal variation in energy usage was handled by projecting sales separately for each month. This approach does not allow for any correlation between sales of energy in successive months. The annual peak, in a majority of the years in the historical record, occurred in the same month each year. Correlation of the system peak demand with class energy sales for this month gave a peak demand model of the form

$$DP_i = a + b.C_i + c.R_i + d.I_i$$

where DP_i is the peak demand in year i, C_i, R_i and I_i are commercial, domestic and industrial sales in the peak month of year i; a, b, c, d are constants to be determined. This model assumes that the components always contribute to the annual peak in the ratio of their respective energy usages, regardless of the time of day when it occurs or of its magnitude. For example consider two successive annual peaks, the first occurs at 9 a.m. on the peak demand day and all components are contributing, the second occurs at 6 p.m. when commercial and most industrial load is approaching zero; clearly the components do not contribute to the peak demand in the same proportions but the model assumes they do. For this reason it is not valid to use a correlation between maximum demand and energy usage unless it is known that all peaks occur at the same time of the day.

Hooke (17), using the scheme outlined above, quotes errors in the peak demand estimates, determined after the true values of the economic variables were known, as not greater than 3% with a mean of 1%. When forecasted values of the economic variables are used the errors are likely to be greater. In actual forecasting over a lead time of 10 months, an error of about 2% was achieved, rising to 8-9% for a 46 month lead time. The allowance for error (uncertainty) was apparently decided from experience and was not quoted. It might be expected to be greater than the achieved errors quoted and as such it is unlikely that the results would be within the desirable 3 - 4% capacity margin (see chapter 3).

Godard (12) used a similar approach to Hooke for industrial sales but forecast domestic consumer numbers with a formula of the form;

$$n_i = ((FRB_i/100)\ (S_i + S_{i-1})/200) \times (1.1658 - 0.0324\ (i - 1946)) \pm E$$

where E represents unusual changes, n_i is the number of consumers in year i, S_i the housing starts in that year and FRB_i the Federal

Reserve Board index of production, corrected for the particular region. The peak month load factor, after correction for seasonal variation and changes in the length of the working week, was found historically to be nearly constant at 55%. Annual peak demand was forecast by combining the energy usage and load factor forecasts. Achieved forecasting errors over five year lead times varied from 0.2% to 4.3%. Again no uncertainty figures were given but from the quoted achieved errors are likely to be 5% or greater. It is not possible to determine whether the method would be satisfactory in New Zealand without experimentation but if the uncertainty had a similar magnitude in New Zealand it would not meet the criterion of desirable accuracy.

The "market research" approach is an attempt to detect changes in the nature of the load and make appropriate allowances in the forecasts. To achieve accurate forecasts of the demand accurate forecasts of the constituent quantities are required. Confidence in the final demand forecast is determined by the confidence in the constituent forecasts, weighted according to their relative contributions to the total demand. Electric energy enters indirectly into many industrial and commercial processes. The principal difficulty with the "market research" approach is the determination of the relationship between sales of electric energy and the output of these processes. Forecasts of product sales are, in many cases, based on subjective analyses of the market for goods, etc (96). Consequently it is difficult to obtain quantitative estimates of the uncertainty in the resultant demand forecasts.

4.2.3 Methods applicable to distribution system planning

The planning of distribution systems requires a knowledge of the characteristics of the load (64). A number of methods of obtaining these from load survey data are available (22, 65, 66). Recently attempts have been made to determine the peak load on a distribution transformer in terms of the appliances it serves (25, 67).

Plienes and Zajic (25) developed an approach from which the probability model of chapter 2 was developed. The probability of an appliance being in the "on-state" is p, and in the "off-state" q, at each instant in time. Assume m_i appliances of the i th type are present in the region served by the distribution transformer. The half hour integrated demand for an operating i th type appliance is denoted by k_i. For a collection of appliances (e.g. in several homes) the mean and variance of the distribution of demand are (from 25))

$$\bar{D} = \sum_{i=1}^{M} k_i \, m_i \, p_i$$

$$\mathrm{var}\ (D) = \sum_{i=1}^{M} k_i^2 \, m_i \, p_i \qquad (4.7)$$

This represents the average maximum demand for a given interval in any day in a selected period. A number of days, N, in this period have similar characteristics (e.g. weather). Equation 4.7 can then be written

$$\bar{D} = \left(\sum_{j=1}^{N} \hat{D}_j \right) / N$$

$$\mathrm{var}\ (D) = \left(\sum_{j=1}^{N} (\hat{D}_j - \bar{D})^2 \right) / N \qquad (4.8)$$

where $\hat{D}_j$ is the maximum demand on day j. The distribution, by the central limit theorem, may be shown to be normal and confidence limits can be established. The probabilities p_i and q_i can be determined for existing appliances via equations 4.7 and 4.8, after k_i has been evaluated, e.g. from appliance nameplates, and m_i determined from a survey etc. Actual tests showed the method was able to estimate demands in the 0 - 50 kVA range to within $\pm$ 6 kVA at the 95% confidence level.

Even the most comprehensive load survey does not examine every possible load condition. Reps (67) attempted to overcome this deficiency by developing a model of the load on a distribution transformer independently of survey data. A small amount of survey data is used to verify the model, but not to build it. The load curve shape on the peak load day is determined by the amount of connected load and its time pattern of use. Artificial appliance usage curves for the peak load day were combined using Monte Carlo gaming techniques to form a simulated daily load curve for a single transformer. Only consumer numbers and the percent saturation of major appliances need be known in addition to the usage curves for any particular load to be simulated. Miller (22) employed a similar combining technique to simulate transformer daily load curves from individual consumer daily load curves, these being obtained from a comprehensive load survey (99).

Both of these comprehensive load models recognize that the demand is a probabilistic process. That of Plienes and Zajic assumes that p_i (and hence q_i) are the same regardless of the time when the peak occurs. This is not necessarily valid, particularly if peaks occur in both the morning and the evening. Reps' method requires considerable experience of appliance usage patterns to obtain realistic results. Both methods, as published, are only suitable for loads which are not changing significantly, i.e. little growth occurs and the same appliances are likely to be in use. Their application to long term forecasting would require knowledge of future consumer numbers and the number and type of appliances available. If they are to be applied to days other than peak load and potentially peak load days a knowledge of how appliance usage varies with changes in the relevant variables, e.g. weather, is required. Both methods would appear to have considerable potential for further development.

4.3 Weather - load models

A number of appliances are used to modify the climate inside buildings (e.g. air conditioners, radiators, etc). Consequently there is a seasonal variation in load in response to seasonal changes in climate. Extremes of cold (or heat in some power systems (30)) in any one year tend to boost the apparent peak demand growth rate. Correction of the historical demand to "standard" weather conditions permits the true growth rate to be determined (39, 59). Changes in the daily load pattern may be anticipated from short range weather forecasts and knowledge of the response of the demand to changes in the weather.

The studies of the weather-demand relationship that have been made, e.g. (14, 24, 30, 56, 87), divide the load into two components;

(a) a base load of a magnitude which is determined empirically as a function of the time of the year and is considered constant for periods of one week to one month, and

(b) a weather sensitive component.

Which weather variables affect the demand depends on the load composition.

One method of forecasting the demand, for generation scheduling, weighted the forecasted weather variables (temperature, cloudiness and wind velocity) and applied the sum of the weights as a percentage modification to a base load (24). The weights (which were determined by a trial and error procedure from historical data) were those which gave a constant base load for each hour of the midweek days of a particular week and also for several consecutive Saturdays and Sundays.

Over lead times of 6-8 hours the achieved forecasting errors were about 0-2%. The accuracy required of the weather forecasts was within $5^{o}F$ for temperature and to distinguish between various cloud categories (e.g. Fair - scattered clouds thin and high, or, Fair - thick haze). These tolerance limits were practical in this case as the weather sensitive load formed only 10-15% of the total. A 20% error in this component reduces to about 2% in the overall forecast. The major source of uncertainty, in this case, stemmed from variations in the largely industrial base load. This method has assumed that the weather weighting is the same at all times of the day. Such an assumption may be troublesome in situations, e.g. New Zealand, where the weather sensitive load approaches 50% of the total load. Varying sensitivity may arise because certain appliances are not greatly used at some times of the day, e.g. electric radiators while people are in bed.

In a more detailed study Davies (14) attempted to define the demand-weather relationship using regression analysis. The weather variables were redefined to represent the effect of the weather on the person. Specifically these were

(a) "Effective" temperature, denoted by X_1

(b) Cooling effect of the wind, " " X_2

(c) Illumination index, " " X_3

(d) Rate of precipitation, " " X_4

The relationship was determined from a year's data, after elimination of a base load, b_o, a smoothed seasonal trend, $F(t)$, and a day of week component δ. The weather sensitive component was assumed to be the sum of contributions from each variable. The linear regression equation took the form;

$$D = b_o + F(t) + \delta + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 \qquad (4.9)$$

The seasonal term was expressed as the sum of orthogonal polynomials up to the sixth order;

$$F(t) = \sum_{i=1}^{6} a_i P_i$$

and assumed to a step function with step length of one week. Here P_i denotes the i th orthogonal polynomial. The day of week variation was expressed in terms of dummy (0,1) variables;

$$\delta' = \sum_{j=1}^{5} \delta_j Z_j$$

where Z_j = 1 on day j, 0 on all other days, and

δ_j = correction on day j.

The quantities of interest, i.e. the δ_j and coefficients b_1, b_2, b_3, b_4, were determined in the subsequent regression analysis.

The restriction assumption of linearity introduced in equation 4.9 was avoided by using a graphical non-linear regression analysis technique. Equation 4.9 was modified to

$$F(t) + \delta = D - g_1(X_1) - g_2(X_2) - g_3(X_3) - g_4(X_4)$$

The quantity on the left may be considered to be the "no-weather" load. Initial estimates of this and of approximations to the g_i were successively modified until no further improvement could be made.

The effective temperature is a measure of the internal temperature of buildings. The thermal capacity of the structure means that internal temperature changes lag external ones. A thermal time constant of 22 hours was established by a correlation analysis. The cooling effect of the wind stems from increased ventilation and infiltration of air. An empirical formula for this in terms of the ambient temperature and wind velocity was developed. The illumination index was found by applying reduction factors for different cloud types to ideal illumination values, these being obtained as a function of the sun's elevation.

Forecasts are made by extrapolating the base load into the future and applying adjustments for the day of the week and the forecasted weather. Simulated forecasts made over a 24 hour lead time using actual weather reports resulted in an RMS error of 1.4%. This represents the amount of variation of demand which is not explained by the model. The quoted forecasts were made only for the peak demand times of the day, i.e. 0900, 1200 and 1800 hours in this case (in England), and required separate analyses of historical data at each interval.

In a system where, due to air conditioning load, the annual maximum demand occurred in the summer Heinemann et.al. (30) devised a weather variable which was linearly related to the cooling (i.e. air conditioning) load, using a regression analysis. With the cooling component determined from historical data, extrapolation was used to forecast its growth (39). The forecasting strategy, which is, common to several weather-load models, is shown in figure 4.1. The mathematical model was similar to equation 4.9, being

$$D_{peak} = b_o + COOL.f(W) + \sum_{i=1}^{6} b_i Z_i \qquad (4.10)$$

Where D_{peak} = daily peak demand

b_o = summer "base load" in May,

b_i = additional base load in the months June to October

Z_i = 1 in month i, 0 in all other months,

COOL = cooling factor, and

f(w) = weather variable.

The function $f(w)$ combines the weather parameters

(a) dry bulb temperature, DB, and relative humidity, RH, at the time of the system peak, and

(b) the dry bulb temperature history over the last three days.

(rain, wind and cloud cover also influence summer peaks, but their effect on the cooling load was not considered significant). The form of $f(w)$ was:

$$f(w) = k_1 \left[f_1(DB,RH)^2 + k_2 f_1(DB,RH) \right] + k_3 \left[DBB^2 + k_4 DBB \right] \qquad (4.11)$$

where $f_1(DB,RH)$ is a heat content factor found from psychrometric charts, e.g. (68), and

$$DBB = k_5 \sum_{i=1}^{3} (k_{8i}\, DB_{max_i} + k_{9i}\, DB_{min_{i-1}}) \exp(-24i/\tau)$$

where DB_{max_i} = highest hourly temperature on each of the three preceding days,

DB_{min_i} = lowest temperature on the peak day and the two preceding, and

τ = time constant of heat build-up (about 3 days in this case).

Using this model, differences between actual and calculated maximum demands varied from 0.5% to 5.7% with a mean of 2.4%. The forecasts were restricted to the daily peak. Uncertainty in the weather sensitive component may be estimated from its probability density function (1, 57) or from past performance (39) but is not quoted by the authors. The restriction of a linear relation between f(w) and the cooling load makes f(w) considerably more complex than it might otherwise have been. It also means that the relationship is likely to be peculiar to the particular power system and consequently considerable effort would be required before the method could be used in other power systems. It should, however, be feasible to use the same approach to determine a heating (rather than cooling) variable which would have wider application (power systems with summer peaks are relatively rare). Unless the daily peak demands used in the regression analysis occur at about the same time each day throughout the summer (i.e. for 6 months), the load composition and hence the true value of the "base load" parameters, b_0-b_6, will vary between observations. In this case the same model does not properly represent all observations, thus tending to increase the variance of parameter estimates.

Long range prediction with weather-load models is achieved by substitution of long term weather statistics, generally for the worst case situation (59). (It is difficult to make weather forecasts over lead times of several years). Short term forecasts use forecasted weather. An error of $2^{\circ}C$ in a temperature estimate is possible over a 24 hour lead time (69) and this combined with a load change of about 2% per $1^{\circ}C$ (9) giving an "error" of about 4%. Extremes of temperature are forecast less often than they actually occur, e.g. 4% of forecasts specified forecasts more than $1.4^{\circ}C$ below average, but actual occurrence was 29% in a study by Freeman (70).

The work of Davies (14) indicates that, given accurate weather forecasts, a comprehensive weather load model can achieve an accuracy of about 3% at the 95% confidence level. This is near the degree of accuracy considered desirable in chapter 3. However, weather forecasting errors are such that the resultant forecasts require a considerably larger margin for error. Furthermore the occurrence of extreme values of, say, temperature and hence of extreme peaks of demand are extremely difficult to predict.

The weather variables which influence the demand are determined by the type of load and its geographic location. Adequate system-wide forecasts, in New Zealand, would require separate forecasts for, at least, the major load centres as these are widely separated geographically and can experience widely differing weathers simultaneously. The character of the weather sensitive component changes with time of day (e.g. radiators are virtually unused during hours of sleep). Thus a weather-load model should make allowance for this changing sensitivity of the demand.

Of the weather-load models discussed the one due to Davies is the most comprehensive as it is

(a) not restricted to particular days (as is Heinemann's), and

(b) based on demand-weather relationships derived from a physical understanding of the effect the weather has on consumer's usage of appliances.

Though considerable computation is needed to obtain the regression coefficients etc. this need only be performed initially and when a significant change in load composition occurs. The results of the computation may be summarized in graphical or tabular form to enable manual forecasting without the use of extensive (and expensive) computing facilities. Originally restricted to specified intervals of the day this model could be extended to all daily intervals with the aid of modern computing facilities. If this is done either the day-of-week factors must be calculated for each interval, or only one calculated for each day and the load curve shape assumed the same for all week days. This assumption is not realistic, as figure 2.2 showed. Alternatively, a model based on the weekly rather than daily load curve, e.g. (10), would avoid the need for day of week factors entirely and also permit the effect of temperature etc. to vary between days of the week.

Reliance on accurate weather forecasting is the major weakness of the present weather-load models. Quantitative weather forecasts are, at the present time, subject to considerable uncertainty, e.g. Maunder (76) obtained standard errors of about 1.1°C in temperature forecasts up to 24 hours ahead. To reduce the effect of this uncertainty on the demand forecasts the time lag between temperature changes external and internal to buildings should be exploited in the forecasting process. Instead of inserting the forecasted temperatures directly into the model a lagged, or "effective", temperature should be calculated from these forecasted values. These lagged forecasted temperatures, which then reflect the regions temperature history of which some at least is known, are then used in the model. Thus future demands can be anticipated from known information.

4.4 Short term load forecasts for system operation.

Over short lead times (up to a few days) weather-load models have been used for forecasting the demand. These require data not readily available to the system controller, e.g. weather information. The use of digital computers for system control has prompted the development of methods using demand data only (2, 72-75). These assume the weather is constant over the required lead time (for system control this is generally a few hours only) and extrapolate a mathematical model given a knowledge of the immediately past demands.

Farmer (72, 73) has developed a method which utilizes the fact that the basic demand pattern repeats itself every 24 hours. Each daily load curve may be considered a member of an ensemble of time series and the forecasting problem is viewed as the prediction of a non-stationary process, knowing this ensemble. The Karhunen-Loeve expansion is used to represent the M sample daily load curves as a linear combination of orthogonal functions over the period (0,T = 24 hours); i.e.

$$D_j(t) = \sum_{i=1}^{K} a_{ji} \lambda_i^{\frac{1}{2}} \phi_i(t) + e(t) \qquad (4.12)$$

where $j = 1, \ldots, M$, and

$e(t)$ = error in the expansion at time t.

The coefficients a_{ji} , $\lambda_i^{\frac{1}{2}}$ and the orthogonal functions $\phi_i(t)$ are found by minimizing the integrated mean square error equation 4.12, giving (see Appendix B)

$$\int_0^T R(t,s) \phi_i(s)\, ds = \lambda_i \phi_i(t)$$

where $R(t,s) = \underset{j}{E}\left\{ D_j(t)\ D_j(s) \right\}$ (4.13)

and λ_i and $\phi_i(t)$ are the eigenvalues and functions of the covariance function R. Also

$$a_{ji} = \frac{1}{\lambda_i^{\frac{1}{2}}} \int_0^T D_j(t)\,\phi_i(t)\,dt \qquad (4.14)$$

Provided the eigenvalues are ordered, the mean square error involved in expanding the load curves to a finite number of terms K is given by

$$MSE = \int_0^T R(t,t)\,dt - \sum_{i=1}^{K} \lambda_i$$

$$= \sum_{i=K+1}^{\infty} \lambda_i \qquad (4.15)$$

(i.e. the error is equal to the sum of the neglected eigenvalues).

Demand forecasts are made by determining a set of coefficients a_{ji}, $i = 1, \ldots, K$, for $j > M$, from the most recent demand values. This assumes that future load curves are generated by the same process that produced the sample curves. The weather is assumed constant over the intervals in which the demand is measured. Methods for determining the a_{ji} included minimizing the mean square error over the last L demand intervals, as in the results reported in Appendix C, (72, 75) or from a conditional probability approach (73). A demand forecast made for r intervals into the future from the present time T_p is given by

$$\hat{D}_j(t') = \sum_{i=1}^{K} a_{ji}\ \lambda_i^{\frac{1}{2}}\ \phi_i(t') \qquad (4.16)$$

where $t' = (T_p + r) \pmod{T}$, and $j > M$

The process generating the sample changes with time, e.g. as winter approaches a heating component is introduced. Thus for $j \gg M$ the load curves cannot be assumed to be generated by the same process as the sample curves. Therefore it is necessary that the orthogonal functions be recalculated at frequent intervals, generally less than one week (73) though the decision may be made automatically (75). While methods of allowing for seasonal trends and day of week effects have been examined (2, 73), the major problem is the choice of the parameters M and K. Too large a value of M includes load curves which are not members of the same ensemble. Toolarge a value of K includes random features peculiar to individual days.

The RMS forecasting error obtained in on-line system control experiments (5, 73) varied from 5% to about 10% for forecasts made from 15 minutes to 2 hours before the time of the daily peak. Using only one orthogonal function (which is equivalent to the mean of the sample curves)RMS errors of 3.4% over a 24 hour lead time, reducing to 0.7% one hour ahead, have been obtained (2). The results of experimental forecasts made using New Zealand load data are given in Appendix C. These experiments gavehigher values of RMS error, from about 6 to 9% for lead times of 1 to 10 hours. This is to be expected as the New Zealand data exhibits considerably greater percentage variation than do those studied by the other authors mentioned. The results in Appendix C suggest that for the loads studied M = 10 and K = 1 are satisfactory.

During the online trials (73) this method was found to be extremely sensitive to data errors; this was also noticed during the experiments of Appendix C. To overcome this problem Farmer substituted a simpler scaling method which achieved forecasting errors of a similar order (73). In this method the load on the day of prediction is compared with a standard load curve for that day. Each measured load is divided by the corresponding value of the standard curve to yield a "present to standard" ratio. Assuming this ratio remains constant over the desired load time the forecasted load values are obtained by multiplying the standard curve by this ratio. The standard curve may be e.g. the average for that day of the week, obtained from historical data, or the immediately preceding day's load curve.

A histogram of the percentage forecasting error over a lead time of one hour using this scaling method is shown in figure 4.2. The standard load curve was the average of the immediately preceding five daily load curves and a total of 780 forecasts were made at half hourly intervals. The variance of the distribution of forecasting error increases considerably with lead times greater than one hour (approximately double over a five hour lead time).

Generation scheduling requires that sufficient accuracy (in the sense of chapter 3) be achieved over lead times greater than or equal to the "start up" time of the generators. In hydro systems this may be of the order of several minutes, for thermal plant up to twelve hours. It is desirable that the sufficient accuracy of about 3-4% at the 95% confidence level (about the capacity of the largest generator in New Zealand) be maintained at all times of the day to minimize fuel costs. The results obtained (in figure 4.2 and Appendix C) indicate that for lead times of one or more hours these methods (scaling and Karhunen-Loeve expansion) are not sufficiently accurate for generation scheduling purposes. Over very short lead times, depending on the composition of the load they may be sufficiently accurate.

In both methods the future is determined solely from historical demand values. Weather is not included in these models as it is reasoned that it has no effect over lead times up to about 2 hours. This reasoning ignores the value of present (and immediately past) weather as an indicator of likely deviations from the "standard" behaviour of the demand. For short lead times the lagged effect of temperature on the demand means that temperature is a leading indicator of deviations of the demand from a "normal weather" daily load curve. Thus both methods would be improved by incorporating present weather (at least) into forecasts over lead times greater than a few minutes.

The Karhunen-Loeve expansion technique (equations 4.12 to 4.16) is considerably more complex, both mathematically and computationally, than the scaling method. Its use does not result in significantly more accurate forecasts and cannot therefore be justified. Simple scaling methods, modified to consider present weather as suggested above, are recommended. These, because of their simplicity, are particularly suited to manual forecasting applications. In a recent paper Toyoda et.al (37) have in fact proposed a forecasting model along these lines; i.e. using weather information to scale a "standard" demand value and also rate of change information which includes by implication the recent weather history.

4.5 Conclusions

This chapter has discussed examples of the main forecasting methods in use, or proposed, to illustrate the present state of the art of electric energy forecasting. The majority of methods in use give essentially single value forecasts, although confidence limits can be calculated in most cases. At present a forecasting method's "goodness" is measured in terms of its performance, i.e. the magnitude of the difference between actual and forecast demands. In chapter 3 it was shown that a more useful measure of "goodness", more closely related to the reliability and cost of supply, is the amount of uncertainty in the forecast at some specified level of confidence, e.g. 95%. It is not possible, however, to compare methods satisfactorily without experimentation with the actual (or typical) data to be used. Where possible an attempt has been made to obtain an indication of the amount of uncertainty which might be expected in the forecasts. Of the methods actually tried on New Zealand data none can be called satisfactory with respect to the desirable accuracy. In the following paragraphs the current state of the forecasting art is summarized.

Long term forecasting relies heavily on trend extrapolation. The extrapolations may be modified on an ad hoc basis to include local knowledge. The trend models used are selected, by professional judgement, to suit the particular system. The accuracy which can be achieved is limited by the variability of historical data. As only a small number of observations are usually available estimates of model parameters are subject to considerable uncertainty. Commonly an allowance for error is established from experience.

Models of consumers' appliance usage behaviour have been used for forecasting demands on distribution transformers. These models require a considerable amount of data, which is not readily available, for their implementation. Their use has been restricted to estimating annual maximum demands. Appliance usage has not been determined for circumstances other than possible peak load days.

Weather-load models have been used for short term forecasting and for estimating uncertainty in long term demand forecasts and "correcting" historical data to "standard" weather conditions. Regression analysis has been used to determine quantitatively the influence of weather (particularly temperature, cloud cover, humidity and rainfall) on the demand at specified times of the day. Which weather variables are most significant depends on the composition of the particular load. Most models have been derived empirically but the variant due to Davies has been based on an analysis of the physical mechanism of the effect of weather on the consumer. It is the most comprehensive of those in use and, provided the necessary data can be obtained, appears to be the "best" in the sense of the desirable accuracy. Experimentation is needed before a categorical statement as to which method is "best" can be made. All these models treat the demand at each interval of the day separately. In fact there is considerable correlation between consumers' usage of appliances (and hence demand) in successive daily intervals. The time lags which exist between changes in, e.g. temperature, and demand should be exploited to enable forecasts to use known information more extensively, rather than relying entirely on weather forecasts.

The short term forecasting methods which utilize demand data only are essentially forms of trend extrapolation. Experiments indicate that the two methods discussed are not sufficiently accurate, in their present form, over lead times of one or more hours. It is desirable that they be modified to enable demands to be anticipated by, e.g., including weather information.

Table 4.1

Commonly used trend curves.

Straight line	$D_t = a + b.x_t + e_t$
Parabola	$D_t = a + b.x_t + c.x_t^2 + e_t$
'S' curve	$D_t = a + b.x_t + c.x_t^2 + d.x_t^3 + e_t$
Exponential	$D_t = a.\exp(\ b.x_t + e_t\ )$
Modified exponential	$D_t = a.\exp(\ b.x_t + c.x_t^2 + e_t\ )$
Logistic	$D_t = (\ 1/(\ a + b.\exp(\ c.x_t\ ))) + e_t$
Gompertz	$D_t = (\ 1/\log_e(\ a + b.\exp(\ c.x_t\ ))) + e_t$

where D_t = actual demand at time t,

x_t = value of the independent variable (usually time) at time t,

a,b,c,d = parameters to be determined by e.g. Least Squares, and

e_t = 'error' between model and actual demands, distributed with known variance about a mean of zero.

TABLE 4.2

NEW ZEALAND ANNUAL MAXIMUM DEMAND FORECASTS

YEAR	ACTUAL MW	Forecast Demand in MW (Percent uncertainty at 95% level)									
		Lead time in years									
		1		2		3		4		5	
1949	555										
1950	577	587	(6.5)								
"	616	636	(5.5)	635	(7.7)						
"	668	668	(5.3)	695	(6.5)	689	(9.4)				
"	681	712	(4.7)	719	(6.2)	761	(7.9)	749	(11.4)		
"	776	732	(5.5)	763	(5.6)	776	(7.6)	837	(9.6)	816	(13.8)
1955	878	817	(6.1)	771	(6.5)	817	(6.8)	837	(9.2)	924	(11.6)
"	912	934	(7.3)	878	(7.3)	811	(7.9)	875	(8.2)	904	(11.1)
"	956	1016	(7.1)	1026	(8.7)	945	(8.8)	851	(9.6)	937	(9.9)
"	1075	1058	(7.8)	1121	(8.4)	1132	(10.5)	1018	(10.7)	891	(11.5)
"	1069	1178	(7.2)	1146	(9.2)	1241	(10.2)	1252	(12.8)	1097	(13.0)
1960	1277	1176	(8.1)	1295	(8.5)	1243	(11.2)	1379	(12.5)	1391	(15.5)
"	1374	1329	(9.4)	1249	(9.7)	1428	(10.4)	1348	(13.7)	1538	(15.1)
"	1554	1473	(9.6)	1434	(11.2)	1323	(11.8)	1579	(12.6)	1464	(16.6)
"	1584	1671	(10.1)	1603	(11.4)	1548	(13.6)	1398	(14.3)	1752	(15.3)
"	1810	1809	(9.1)	1844	(12.0)	1746	(13.9)	1670	(16.6)	1475	(17.4)
1965	1976	2032	(8.8)	2004	(10.8)	2039	(14.6)	1904	(17.0)	1802	(20.2)
"	2176	2196	(8.5)	2280	(10.4)	2228	(13.2)	2260	(17.9)	2077	(20.6)
"	2415	2390	(8.0)	2435	(10.0)	2572	(12.7)	2487	(16.1)	2512	(21.8)
		2653	(8.2)	2629	(9.6)	2706	(12.2)	2917	(15.5)	2785	(19.6)
				2931	(9.7)	2895	(11.6)	3015	(14.9)	3326	(18.8)
						3241	(11.8)	3190	(14.2)	3367	(18.1)
								3588	(14.4)	3518	(17.2)
										3877	(17.4)

TABLE 4.3

NEW ZEALAND ANNUAL ENERGY USAGE FORECASTS

YEAR	ACTUAL GWh	Forecast Energy Usage in GWh (Percent uncertainty at 95% level)				
		Lead time in years				
		1	2	3	4	5
1949	2669					
1950	2871	2756 (4.9)				
"	2942	3051 (3.0)	2889 (5.8)			
"	3299	3166 (2.9)	3273 (3.5)	3022 (7.0)		
"	3397	3502 (4.0)	3361 (3.0)	3516 (4.3)	3153 (8.5)	
"	3847	3696 (4.0)	3778 (4.7)	3567 (3.0)	3784 (5.2)	3282 (10.3)
1955	4197	4092 (4.6)	3971 (4.7)	4086 (5.7)	3787 (3.1)	4080 (6.2)
"	4526	4550 (4.8)	4444 (5.4)	4275 (5.7)	4430 (7.0)	4019 (3.1)
"	4793	4761 (5.2)	4998 (5.7)	4841 (6.6)	4610 (6.9)	4814 (8.4)
"	5416	5327 (5.2)	5123 (5.3)	5513 (7.0)	5287 (8.0)	4980 (8.3)
"	5444	5947 (5.1)	5837 (6.1)	5512 (5.4)	6106 (8.4)	5790 (9.6)
1960	6124	6069 (7.0)	6586 (6.1)	6413 (7.5)	5931 (5.5)	6789 (10.2)
"	6576	6480 (6.2)	6544 (8.3)	7322 (7.4)	7064 (9.1)	6381 (5.6)
"	7155	7026 (6.2)	6922 (7.3)	7055 (10.1)	8172 (9.0)	7803 (10.9)
"	7676	7566 (5.9)	7532 (7.4)	7377 (8.9)	7602 (12.3)	9157 (10.8)
"	8725	8308 (5.1)	8084 (6.9)	8063 (8.9)	7843 (10.8)	8188 (14.9)
1965	9455	9433 (5.4)	8985 (6.0)	8619 (8.4)	8620 (10.9)	8320 (13.0)
"	10260	10470 (5.5)	10401 (6.5)	9721 (7.3)	9170 (10.3)	9201 (13.2)
"	10993	11444 (5.4)	11626 (6.5)	11506 (7.8)	10522 (8.8)	9736 (12.4)
		12348 (5.8)	12692 (6.4)	12959 (7.9)	12768 (9.5)	11392 (10.7)
			13639 (6.8)	14123 (7.8)	14500 (9.6)	14214 (11.5)
				15103 (8.3)	15765 (9.5)	16288 (11.6)
					16766 (10.1)	17656 (11.4)
						18661 (12.2)

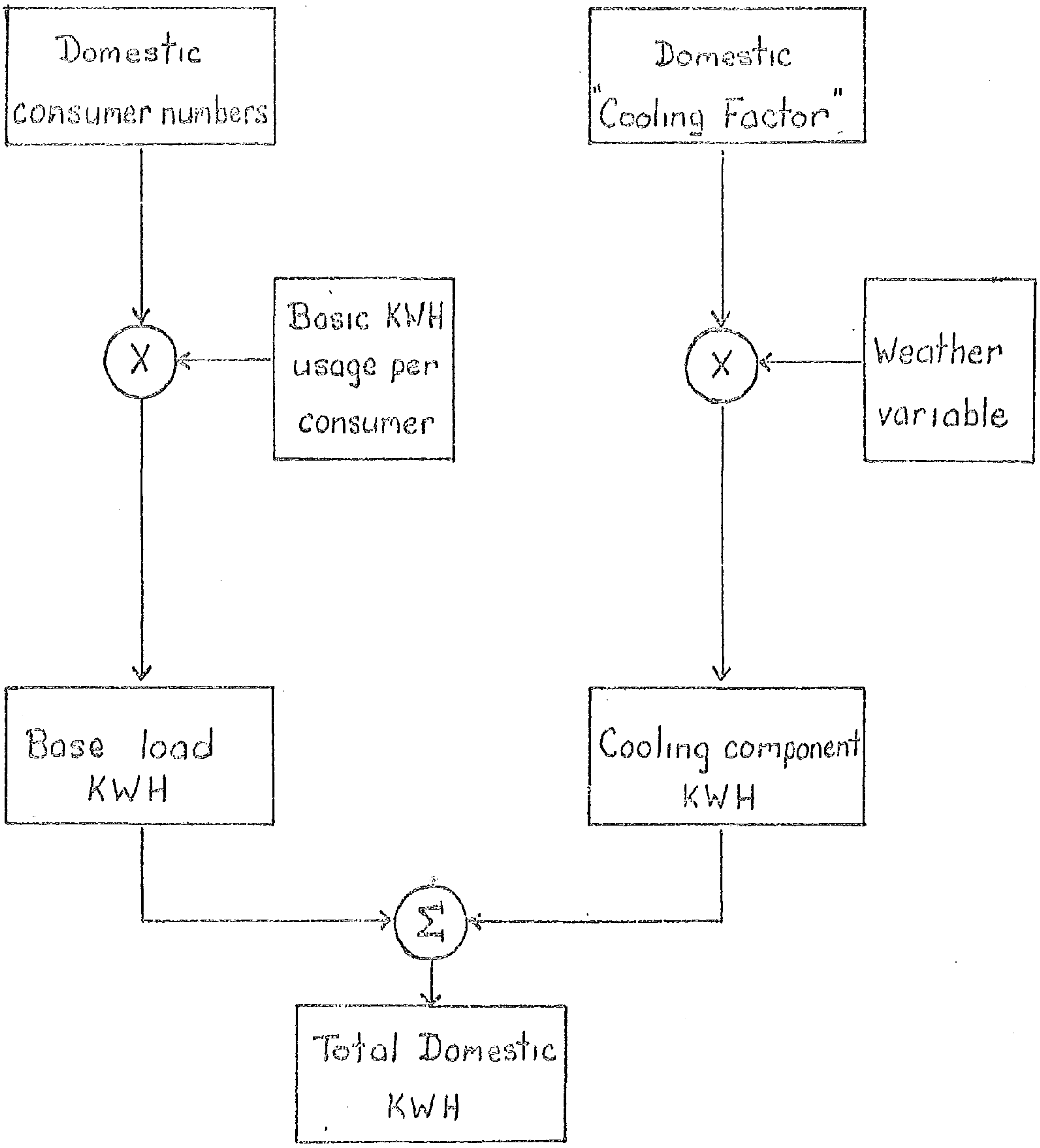


Figure 4.1 Typical forecasting strategy using a weather-load model (from Latham et. al.)

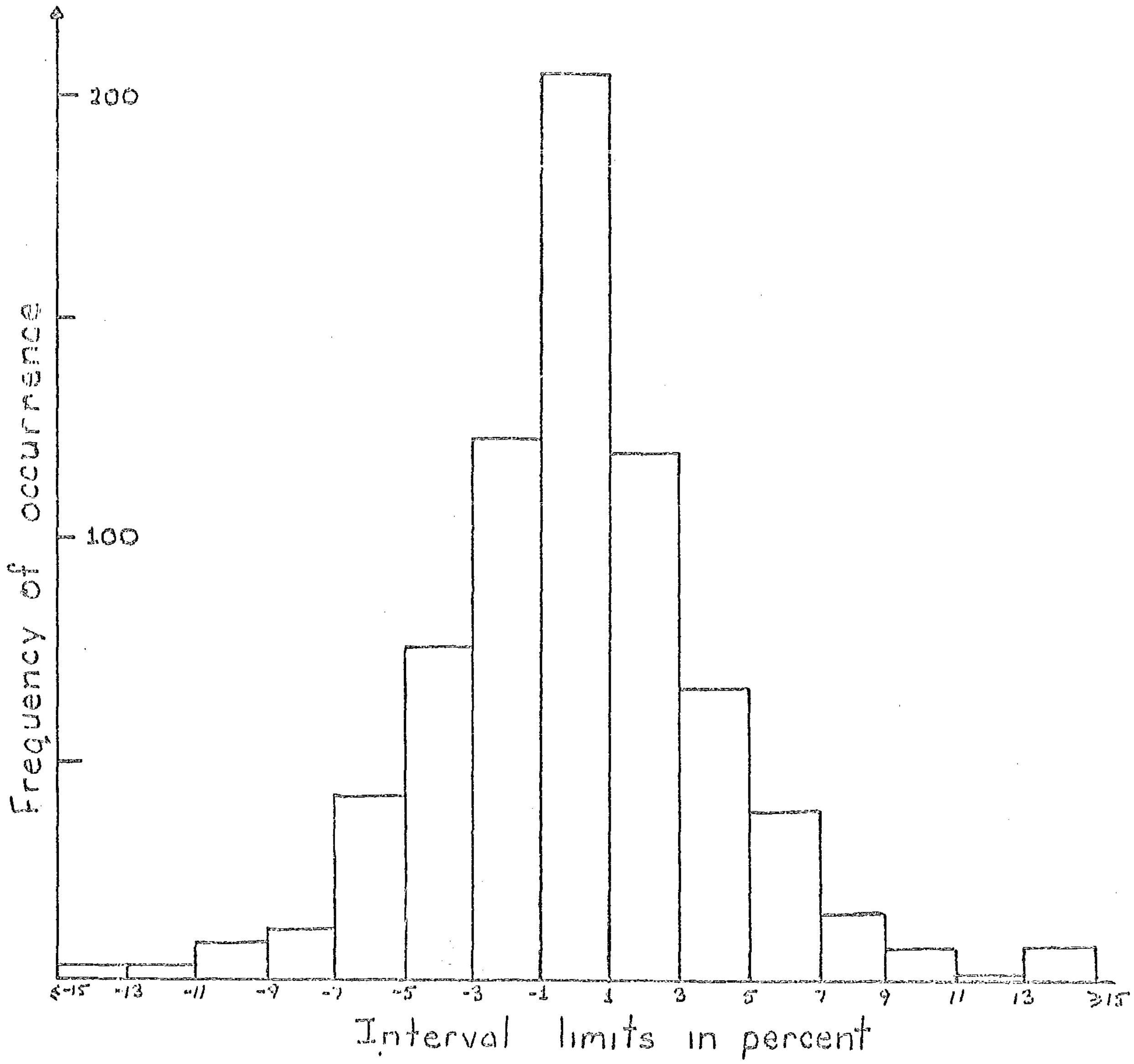


Figure 4.2 Error histogram for scaling method. Standard daily load curve was moving average over 5 preceding days.

CHAPTER 5

THE LOAD GROWTH PROCESS

5.1 Introduction

The load growth process increases the capacity of the load to use electric energy. It determines the annual increases in both maximum demand and energy usage. The installation of new equipment to meet these increases is planned with the aid of long term forecasts, i.e. forecasts over lead times of several years, of these quantities.

The commonly used forecasting procedures, as Chapter 4 showed, are based on the extrapolation of a mathematical time series fitted to the historical data; i.e. trend extrapolation. Alternatively the growth rate may be treated as a random variable distributed in some way about a long term mean value (50, 77).

A disadvantage of time series extrapolations is their inability to foresee deviations from the trend which may often be anticipated from other information. Such deviations result from factors, not present in the historical record; e.g. planned increases in industrialization. The forecaster may use his "judgement" to modify the extrapolation if, in his opinion, the results of market surveys, local knowledge etc, suggest a deviation is likely (17, 11). This introduces an element of anticipation into the forecasting procedure which is not present in pure extrapolation. Also correlation with economic variables may be introduced at this stage (17).

An ability to modify an extrapolation implies some knowledge, however meagre, of the contribution of variables other than time to the growth process. The nature and extent of these contributions are determined by a set of rules; these form the structure of the load growth process. Modification therefore utilizes a structural model of the growth process though this may only be a qualitative model; as used by New (11). A structural model represents the actual set of rules by a known, usually simplified, set such that the growth behaviour of the model and the actual load are equivalent.

In this chapter a load growth structure is established on which empirical models can be built. This structure is inferred from the basic facts outlined in Chapter 2; i.e. that the number of consumers and the average energy usage in each class is growing with time. A lack of basic data precludes conclusive proof of several aspects of the resulting structure which, as a consequence, is mainly of an experimental nature.

The time series and economic correlation models used at present are static in nature, i.e. there is no reference to any time lags which may exist between a "stimulus" and the resultant demand - a "response". This means that separate forecasts of the contributory variables must be made before a demand forecast can be obtained. These can be as difficult to obtain as a direct demand forecast (29). Where time lags exist they have been incorporated into the inferred growth structure to enable dynamic models of load growth. Dynamic models are particularly useful when assessing whether a particular control action, e.g. tariff modification, will affect future energy requirements in the desired way.

Attention is confined to annual growth of energy usage, as distinct from maximum demand. This is because the maximum demand in any year is determined by appliance usage patterns, and is thus a short term phenomenon. The approach adopted is as follows. The determinants of growth of consumer numbers, in both the domestic and non-domestic sectors, are considered. Subsequently the way in which a consumer's energy requirements vary in the long term are discussed. The order and velocity of accumulation of household appliances forms the basis of the domestic consumer's growth process. Non-domestic energy usage is shown to be determined by the volumes of production and sales of goods and services.

5.2 Load growth for a class of consumers

Each class of consumer is distinguished by a characteristic primary use for electricity; see Chapter 2. Because of this, the contribution from each class to the total system load growth may be assumed to be governed by similar factors. Thus only one long term load growth model is needed for each class of consumer, rather than one for each individual consumer.

The electricity requirements of a class of consumers is a function of the number of consumers in the class and their individual requirements. Let $N_t^{(k)}$ denote the number of consumers in class k during the time interval $(t, t + \Delta t)$, denoted by t, and w_{it} the energy usage of the i th consumer. The amount of energy used by class k in interval t is

$$W_t^{(k)} = \sum_{i=1}^{N_t^{(k)}} w_{it} \qquad (5.1)$$

Now let $\Delta N_t^{(k)}$ denote the number of new consumers formed in interval t and w_{jt} the energy usage of the j th new consumer, $j = 1 \ldots, N_t^{(k)}$. Let Δw_{it} denote the change in energy usage of the $N_t^{(k)}$ consumers already in existence. The energy usage in interval (t + 1) of class k is then

$$W_{t+1}^{(k)} = W_t^{(k)} + \sum_{j=1}^{\Delta N_t^{(k)}} w_{jt} + \sum_{i=1}^{N_t^{(k)}} \Delta w_{it} \qquad (5.2)$$

The amount of energy which the power system must supply in interval t is simply

$$W_t = \sum_{k=1}^{C} W_t^{(k)} \qquad (5.3)$$

where C = number of consumer classes distinguished.

The growth of system load in each time interval is determined by the processes which determine $\Delta N_t^{(k)}$ and Δw_{it} for each consumer. These growth processes are discussed in the following sections for the two classes; domestic and non-domestic.

5.3 The growth of domestic energy requirements.

5.3.1 The growth of domestic consumer numbers : basic model.

The regional population is distributed, according to some scheme, amongst the available dwellings. Each group of people in a dwelling forms a domestic consumer. In any time interval t the i th consumer has a population denoted by P_{it}. Dropping the superscript (k) for simplicity, the number of domestic consumers is N_t. The total population in the region in interval t is

$$P_t = \sum_{i=1}^{N_t} P_{it} \qquad (5.4)$$

The total number of consumers increases as the total number of dwellings. Hence the process governing the increase in consumer members is closely related to that governing the increase in dwelling numbers. The basis of this process is the increase in population.

In figure 5.1 the growth process for domestic consumer numbers is illustrated. As the regional population grows young people leaving home, to work in different parts of the region, to marry etc, and immigrants to the region form a "latent" population. As this "latent" population forms into groups which then occupy dwellings new consumers are formed. However a consumer can only form if a dwelling is available. Dwellings become available as a result of new construction and existing dwellings becoming vacant. The latter occurs when the occupants emigrate, physically die, or become members of some other existing consumer, e.g. old people entering homes. Hence the rate of increase of consumer numbers is restricted by the rate at which dwellings become available.

The latent population in interval t is composed of people born x time intervals, i.e. in interval $t - x$, previously, $0 < x < \infty$, and immigrants, I_t in number. Denote the number of births in interval $(t - x)$ by ΔP_{t-x}. Not all people leave home at the same age; let r_x, $0 \leqslant r_x \leqslant 1$ denote the fraction of people of age x leaving home in interval t. The size of the latent population in interval t is then

$$LP_t = I_t + \sum_{x=0}^{\infty} r_x \cdot \Delta P_{t-x} \qquad (5.5)$$

These people form $\Delta N_t'$ consumers where

$$N_t' = \psi_1 (PL_t) \qquad (5.6)$$

During interval t E_t people emigrate. People of age y die, go into homes etc during this interval. Let ΔP_{t-y} be the number of births y intervals previously and s_y, $0 \leqslant s_y \leqslant 1$ be the fraction of these people who die in interval t. The number of people no longer distributed amongst the available dwellings is

$$DP_t = E_t + \sum_{y=0}^{\infty} s_y \cdot \Delta P_{t-y} \qquad (5.7)$$

and the number of dwellings vacated is

$$N_t'' = \psi_2 (DP_t) \qquad (5.8)$$

Hence the net increase in consumer numbers in interval t is

$$\Delta N_t = \Delta N_t' - \Delta N_t'' \qquad (5.9)$$

$$= \psi_1(LP_t) - \psi_2 \ (DP_t)$$

The number of domestic consumers in interval (t + 1) is then

$$N_{t+1} = N_t + \Delta N_t \qquad (5.10)$$

5.3.2 A simple illustration of the basic process.

The functions ψ_1 (LP_t) and ψ_2 (DP_t) must be determined for the particular region. Similarly the weights r_x, s_y are characteristics of the regional population. If consumers only formed after marriages and dwellings only became vacant on the death of the occupants then r_x and s_y, in New Zealand, would have the form shown in figure 5.2, based on tables in the New Zealand Year Book (33).

For the purpose of illustrating the model of consumer number growth in the previous section assume the following, unlikely, situation in New Zealand. There is no immigration or emigration. The numbers of males and females in the population are always equal. All people marry at the age of twenty and form consumers. At the age of 70 all people die thus vacating their dwellings. In this situation

$$N_t' = 0.5 \ \Delta P_{t-20}$$

$$N_t'' = 0.5 \ \Delta P_{t-70}$$

The net increase in consumer numbers is then

$$\Delta N_t = 0.5 (\Delta P_{t-20} - \Delta P_{t-70}) \qquad (5.11)$$

In figure 5.3 ΔN_t calculated from equation 5.11 using published increases in population (33), and making the assumptions outlined above, is plotted together with the actual increase in consumer numbers for the years 1960 - 1967.

This example makes no allowance for immigration and emigration or changes in average population per consumer. Neglecting the latter assumes the functions ψ_1 and ψ_2 to be constant in time.

Much of the discrepancy in magnitude between the two plots in figure 5.3 is the result of people living longer than 70 years and thus occupying homes for longer; figure 5.2(a) shows that about 50% of all persons live more than 70 years. Also the example does not allow for the demand for flat type accommodation from groups of young people (less than 20 years old), e.g. students, or for people marrying when younger than 20. The discrepancy in form, i.e. actual growth is decreasing while "model" growth is increasing, can be explained by changes in the amount of finance available for home building. From about 1958 considerable assistance was made available, by the Government, in the form of "Family benefit capitalization" (an advance on moneys due) (33). The value of this assistance has since been eroded by inflation. Since about 1967, when a mild recession occurred, loans for housing finance have been increasingly hard to obtain and this, together with a high rate of inflation, has produced the pronounced drop in the growth of domestic consumer numbers.

5.3.3 Formulation of growth of a consumer's energy usage

Growth in energy usage occurs if more appliances are bought and, or, existing ones are given more use. Domestic appliances belong to that class of goods known to economists as consumer durables. Pyatt (13) has developed modellings of the accumulation, by households, of durable goods of all kinds. Much of this work can be applied to the accumulation of electrical appliances and hence to modelling the growth of a consumer's energy usage.

Assume consumers can only choose to own appliances in a set B. A consumer need not own all the appliances in B, which may be made large enough to permit the ownership of several appliances of one type, e.g. several radiators or radios. Let β_{it} denote a set of appliances actually owned by consumer i in the interval $(t, t+\Delta t)$; $\beta_{it} \subset B$, also from Chapter 2, $\beta_{it} \subset A$. There may, but not necessarily, be as many different subsets β_{it} as there are domestic consumers in the region. Let $\max(c_j)$ denote the maximum rate of energy usage (demand) of a member j, of B. The maximum amount of energy which consumer i can use in the interval t is then

$$\max(w_{it}) = \sum_{j \in \beta_{it}} \max(c_j) \; \Delta t \qquad (5.12)$$

The actual amount of energy used is a function of usage given each member of β_{it}. Let u_{ij} be a usage factor, $0 \leqslant u_{ij} \leqslant 1$, such that the actual usage in interval t is given by

$$w_{it} = \sum_{j \in \beta_{it}} u_{ij} \max (c_j) \Delta t \qquad (5.13)$$

Assume that u_{ij}, $j \in \beta_{it}$, does not change from one interval to the next. Growth of a consumer's energy usage now occurs only when additional appliances are purchased, and used; i.e. $u_{ij} > 0$ for $j \in B$, $j \notin \beta_{it}$. Assume also that worn out appliances are immediately replaced so that the subset β_{it} never decreases with time. This means that original purchases of members of B must occur before growth of energy usage can occur.

The probability that consumer i buys an appliance j in the interval $(t, t + \Delta t)$ is then the joint probability that i chooses to buy j next from the alternatives available, and that a purchase is actually made in interval t and that the appliance is not already owned. Let J_j' be the event that j is purchased next, I_t the event that a purchase is made in interval t and L_β the event that a subset β ,$\beta \subset B$ is owned. Hence;

p (consumer i buys j in interval t)

$$= \sum_{\forall \beta,\ j \notin \beta} p\,(J_j \cap I_t \cap L_\beta) \quad \text{for } j \in B \qquad (5.14)$$

The summation is over all subsets β , of which j is not a member.

The quantity being summed in equation 5.14 may also be written,

$$p\,(J_j \cap I_t \cap L_\beta) = p\,(J_j | I_t \cap L_\beta)\; p\,(I_t | L_\beta)\; p\,(L_\beta)$$

$$\forall\ j \text{ and } \beta\ ,\quad j \notin \beta \subset B \qquad (5.15)$$

The expected demand from consumer i for appliances j in the interval $(t,\ t + \Delta t)$ is then

$$E \begin{pmatrix} \text{consumer i buys} \\ \text{j in } (t,\ t+\Delta t) \end{pmatrix} = \sum_{\substack{\forall\ j \notin \beta \\ \beta \subset B}} p\,(J_j | I_t \cap L_\beta)\, p\,(I_t | L_\beta)\; p\,(L_\beta) \qquad (5.16)$$

Let m(B) denote the number of members in B; hence there are $(2^{m(B)}-1)$ possible subsets β. There are a total of $m(B)\ x(2^{m(B)}-1)$ terms of the form $p(J_j | I_t \cap L_\beta)$ possible. These terms are statements of the relative priorities given to ownership of each appliance in B by consumer i. The $(2^{m(B)}-1)$ terms, $p(I_t | L_\beta)$, form a measure of the rate at which the consumer accumulates appliances. The probability that a consumer owns each appliance in B can be determined from the $(2^{m(B)}-1)$ probabilities $p\,(L_\beta)$; i.e.

$$p \begin{pmatrix} \text{consumer owns} \\ \text{j in interval t} \end{pmatrix} = \sum_{\forall \beta\ ,\ j \in \beta \text{ and } j \in B} p\,(L_\beta)$$

$$= OWN_t\,(j) \qquad (5.17)$$

The probabilities $p(L_\beta)$ together with equation 5.17 may be used to determine the level of ownership of appliances in B. Hence if the time path of the $p(L_\beta)$ can be determined then so can the accumulation of appliances. Each member of B has a known maximum demand, $\max(c_j)$. Provided the usage factor u_{ij} can be estimated for each j, $j \notin \beta_{it}$, $j \in B$, e.g. by observation of consumers owning j, then the growth of energy usage by consumer i can be determined.

Now the quantity on the left of equation 5.16 is the expected demand for each appliance in B; the quantity $OWN_t(j), \forall j \in B$, is the probability that each appliance is owned. Let $\dot{OWN}_t(j)$ denote the rate of change of probabilities of ownership. Then

$$\dot{OWN}_t(j) \, \Delta t = E \begin{pmatrix} \text{consumer } i \text{ buys} \\ j \text{ in } (t, t + \Delta t) \\ \end{pmatrix} \quad (5.18)$$

$$\forall j \in B \text{ and } t > 0$$

This holds provided (see Pyatt (13));

(a) the priority elements, $p(J_j | I_t \cap L_\beta)$, remain constant from time zero to time present and

(b) the rates of accumulation of appliances for two subsets β, β^1 with $m(\beta) = m(\beta^1)$ are the same from time zero to time present.

From equation 5.18 the time pattern of movements in appliance ownership may be ascertained, and hence the growth of energy usage. So far nothing has been said about the form of the quantities $p(J_j | I_t \cap L_\beta)$ and $p(I_t | L_\beta)$. This is the subject of the next section.

5.3.4 The determinants of the priority and velocity of appliance accumulation

The priority terms, $p(J_j | I_t \cap L_\beta)$, state the strategy which a consumer would adopt were it actively accumulating goods. Consumers tend to classify appliances into "necessary" and "luxury" items. The strategy is such that "necessary" items are bought before "luxury" items. Clearly, however, what constitutes a "necessary" item can vary between consumers. In this section an attempt is made to identify the factors influential in the formation of the strategy.

Firstly there are physical restraints on the choices a consumer can make. A consumer with an efficient oil fired central heating system is unlikely to purchase electric radiators.

Secondly personal taste is influential because $p(J_j | I_t \cap L_\beta)$ is a statement of consumer preferences. Personal taste dictates the division between "necessary" and "luxury" items and the ordering of appliances in each category; e.g. some consumers may prefer a washing machine before a refrigerator and vice versa.

A third influence is the relative prices of all goods and electrical appliances in particular. The relative prices of appliances influence the $p(J_j | I_t \cap L_\beta)$ to the extent that items of otherwise similar attractiveness are ranked according to relative price. Clearly, however, expensive but "necessary" items, such as electric ranges, will most likely have priority over "luxury" but inexpensive items such as radios.

Pyatt (13) and Speight (28) suggest that the $p(J_j | I_t \cap L_\beta)$ is largely independent of the prices of goods except those items in B which the consumer does not already own. This is reasonable as the $p(J_j | I_t \cap L_\beta)$ makes no statement about preferences for other goods. The prices of other goods may however affect the velocity with which electric appliances are accumulated. While the ordering, or priority, given to the purchase of appliances, and other durable goods, is dependent on relative prices the velocity of accumulation is dependent on absolute prices. Consequently a price change has little effect on the $p(J_j | I_t \cap L_\beta)$ if it does not change the price relativity.

Changes in a consumer's income and wealth (which may be viewed as accumulated purchasing power (13)) have the inverse effect to changes in absolute prices; i.e. an increase in absolute income or wealth has the same effect as a decrease in absolute price. An increase in wealth relative to prices leaves the consumer able to purchase appliances sooner than originally planned. The availability of credit and renting facilities influences the velocity of accumulation; if credit is available then a consumer can finance a purchase out of income and obtain the appliance immediately, rather than wait until sufficient wealth (i.e. savings) has accumulated.

Changes in the price of goods in B, but not yet owned, relative to the price of goods, both durable and perishable, outside B alters the velocity of accumulation; e.g. a relative increase in the price of food reduces the amount of money available for the purchase of appliances and hence slows the velocity of accumulation.

An important factor, not discussed so far, is the number of people and their age distribution forming the consumer. The number of "necessary" appliances in a household, e.g. electric range, refrigerator, water heater, is largely independent of household population. The number of "luxury" items such as portable radiators, radios etc does vary with the household population. As the children in a family grow they become more involved in energy usage, having their own radiators etc in their own rooms when in their late teens. The influence of the size and age distribution of the household population on the priority matrix is largely restricted to the "luxury" items.

5.3.5 Evaluation of consumer load growth

Evaluation of the model of consumer load growth requires knowledge of the energy requirements of all appliances in the set B, the priority terms $p(J_j | I_t \cap L_\beta)$ and accumulation velocity terms $p(I_t | L_\beta)$, together with the usage factors, u_{ij}, for each consumer. While the general form of $p(J_j | I_t \cap L_\beta)$, $p(I_t | L_\beta)$ can be outlined, as discussed in the previous section, a comprehensive evaluation requires a field survey. Such a survey has not been carried out at the time of writing, but see (94) and (96).

The major objectives of a field survey would be to determine, for a specified set of available appliances;

(a) the subsets β_{it} for each consumer surveyed,

(b) the priority given by each consumer to the purchase of appliances not yet owned,

(c) how soon the consumer expects to be able to purchase these appliances, and

(d) the usage given members of β_{it}.

From objectives (b) and (c) the probabilities $p(J_j | I_t \cap L_\beta)$ and $p(I_t | L_\beta)$ respectively may be determined. From (a) and (d) the probabilities of ownership of appliances and the usage given them may be determined. Combined with a knowledge of max $(c_j), \forall j \in B$, the expected growth in a consumer's energy usage may be determined.

5.3.6 The appliance usage factor.

The appliance usage factor is extremely important in the determination of actual energy usage; so far it has been assumed constant. Changes in the population of a household affect the use of owned appliances. This is observed particularly for the "necessary" appliances such as electric ranges and water heaters. A large household population is likely to utilize bathing and laundering facilities to a greater extent than a small one, thus raising hot water consumption and the usage factor of the water heater. A similar effect is observed for ranges. Consequently, the appliance usage factor is seen to be some function of the household population:

$$u_{ij} = g(p_i) \quad , \quad \forall j \in \beta_i \tag{5.20}$$

The exact form of the relationship equation 5.20 must be determined by field measurements on appliances owned by consumers similar in all respects except for population. These have not been done although such a survey was to have been carried out in New Zealand in 1969 (7).

The consumer's wealth or income is not likely to have a significant effect on u_{ij}. A consumer tends to use appliances owned without much regard for operating cost. The usage factor may be modified to the extent that a consumer can cut down on wastage of energy (29).

5.4 Growth of non-domestic energy requirements

The growth of non-domestic energy requirements cannot be treated in the same way as the growth of domestic requirements. Non-domestic consumers exist to supply goods and services demanded by the regional population (and, by implication, domestic consumers). New consumers form only if there is sufficient demand for a particular good, or service, in excess of that which can be met by existing consumers. The demand is considered "sufficient" if the consumer can make a profit on sales (28, 19). Consequently new consumers are not generated by existing consumers as in the domestic case; rather growth of consumer numbers occurs as a result of expansion of the market for goods and services.

The operations carried on by each non-domestic consumer dictate what electrical appliances are owned and hence the consumer's potential energy usage. Because the purchase of appliances represents a cost, the consumers only own those necessary for the particular operation. Purchases are generally made at the time the consumer is formed and there is no subsequent accumulation of appliances as in the domestic case.

5.4.1 Growth of commercial energy usage.

Commercial consumers distribute goods and services to the regional population. The demand for any particular item comes from that portion of the population with a need, or desire, for that item and the ability to pay for it. The whole population contributes to the demand for "necessary" items such as basic foodstuffs, clothing and some household equipment, while for "luxury" items, e.g. recreational equipment, the demand comes from only a fraction of the population.

The population expends its income firstly on "necessary" goods, secondly on "luxury" goods. Consequently the prices of all goods relative to the populations' income has a direct effect on the demand for goods and hence on the volume of sales (13, 28). Each commercial consumer desires to make a profit on sales. For any particular item the sales volume is restricted by population size. The amount of profit made on an item is limited by buyer resistance, competition etc (28). Hence if each consumer is to make a profit some minimum volume of sales is required. This sets an upper limit on the number of commercial consumers, as a function of population size, given that the populations' income is fixed.

There is a limit on the distance people are prepared to travel, and which it is economical to travel, to obtain goods and services. This limit is relatively low for "necessary" items, higher for "luxury" items. The limit can be modified by altering the prices of goods, e.g. supermarkets offer goods at lower prices than corner groceries and hence attract custom from a wider area. As a result population density imposes a lower limit on the number of consumers; this applies mainly to consumers trading in "necessary" items such as foodstuffs.

In any interval of time $(t, t + \Delta t)$ the number of commercial consumers in existence, $N_t^{(3)}$, is some fraction of the regional population, P_t ; i.e.

$$N_t^{(3)} = h_t \cdot P_t \qquad (5.21)$$

where h_t = constant of proportionality such that $0 < h_t < 1$. The bounds on $N_t^{(3)}$ imposed by population density and profitability restrict h_t to a smaller range than (0, 1). The actual range may vary between geographical regions.

As the population grows so does the demand for goods and services. Given a sufficiently large population increase the increase in commercial consumer numbers is:

$$\Delta N_t^{(3)} = N_{t+1}^{(3)} - N_t^{(3)}$$

$$= h_{t+1} \cdot P_{t+1} - h_t P_t \qquad (5.22)$$

where $P_{t+1} = P_t + \Delta P_t$

Provided that there are no changes in

(a) the proportion of the population desiring any particular item,

(b) the age distribution of the population, and

(c) the volume of sales required for a consumer to make a profit, then

$$h_{t+1} \simeq h_t \qquad 0 < h_{t+1},\ h_t < 1 \tag{5.23}$$

If at the same time there is an increase in prices relative to incomes then the increase in the demand for goods and services, and hence the volume of sales, is reduced (28). Fewer new consumers form and

$$h_{t+1} < h_t \tag{5.24}$$

An increase in the number of people who must be served for a consumer to make a sufficient profit has the same effect. If provisos (a) and (b) above do not hold then the parameter h_t may or may not change. A change in (a) implies a change in the tastes of the population, this may merely modify the types of goods and services being offered for sale. A change in (b) may mean completely different demands, e.g. a demand for "mod" clothes, resulting in more consumers forming to handle the changed demand. Such changes are likely to be slow but difficult to anticipate any great time in advance.

Knowledge of the number of commercial consumers does not imply a knowledge of their energy requirements. The number, and type, of appliances owned by each consumer is largely determined by the size of the premises, as shown in Chapter 2; e.g. the larger the premises the more lighting and heating required for a given standard. Size, for a commercial consumer, is dictated to a large extent by the physical volume of sales. To handle a given volume requires a work-force of a certain size. Hence the size of the work-force employed is also a measure of the size of the premises. If β_{it} denotes the set of appliances owned by the i th consumer in interval $(t, t + \Delta t)$ then the number of members of β_{it} are given by

$$m(\beta_{it}) = g(LB_{it}) \qquad i = 1, \ldots, N_t^{(3)} \tag{5.25}$$

where LB_{it} = number of workers employed by i in interval t.

The energy usage of consumer i in interval t is then

$$w_{it} = \sum_{\forall j \in \beta_{it}} u_{ij} \max(c_j) \, \Delta t \tag{5.26}$$

and the total energy requirements of all commercial consumers becomes

$$W_t^{(3)} = \sum_{i=1}^{N_t^{(3)}} w_{it} \tag{5.27}$$

However the size of β_{it} is generally fixed at the time the consumer is formed. It is expensive for a consumer to make any major change of appliance types, hence $\max(c_j)$, $j \in \beta_{it}$, is also fixed at formation. The usage each appliance is given is dictated largely by the hours of work, which in New Zealand are well defined. Consequently the energy usage of existing commercial consumers does not grow to any great extent. Any variations with time are likely to be random, reflecting changes in u_{ij} due to weather, economy drives by management etc.

It is difficult to estimate the energy requirements of consumers not yet formed over lead times greater than the time to erect buildings. If, for example, the volume of sales required for profitability increases, which for a given ΔP_t results in a reduction in $\Delta N_t^{(3)}$ (and h_{t+1}), then the size of new consumers increases. The extent to which the resultant increase in average energy usage compensates for the reduced number of consumers is difficult to estimate. As a consequence extrapolation of historical average energy usages must still be used to determine future energy requirements for this class of consumer.

5.4.2 The growth of industrial energy usage.

Goods, to meet the demand from the population, are obtained from imports and from production within the region. The industrial consumers produce goods for use within the region and additionally for export. Establishment of an industrial consumer requires that a market for the goods to be produced exists and that the "factors of production", i.e. labour and capital, are available in sufficient quantities (19, 20, 28).

A production process, once established, requires inputs of raw material, labour and energy. The contribution of each, measured in some suitable units (usually monetary), to a unit of production is dependent on the particular process, e.g. electronic equipment has a low raw material and high labour content, aluminium smelting has a high raw material and energy but low labour content. The degree of mechanization in a plant alters the relationship between the energy and labour contributions to the final product (32). Of the total energy requirements of industry only a portion is met by electricity, the size of the portion is also dependent on the process; c.f. iron and aluminium ore smelting. Electricity is used extensively for lighting and motive power, and to a lesser extent (because of relatively high costs) for heating.

Once a manufacturing process is in operation the relationship between the inputs for a given amount of production is unlikely to change. This is because a considerable cost is associated with such changes, which in general require the replacement of existing plant by new plant. Hence the inputs are linearly related to the output of a given manufacturing process. Consequently, where electric energy is an input to a process, the amount used in interval t is proportional to the amount of production in interval t; i.e. the electric energy usage of consumer i in class 2 (industrial) is given by

$$\begin{aligned} w_{it}^{(2)} &= \phi\ (\underline{\mathrm{PROD}}_{it}) \\ &= \gamma_i\ \underline{\mathrm{PROD}}_{it} \end{aligned} \qquad (5.28)$$

where $\underline{\mathrm{PROD}}_{it}$ = production in interval t

γ_i = proportionality coefficient

$(\gamma_i \geq 0$ for $\underline{\mathrm{PROD}}_{it} \geq 0)$

The coefficient γ_i is peculiar to each production process. Its value varies with the good(s) being produced and with the degree of mechanization. For example, consider several industrial consumers all producing the same good by different processes. The consumer with the greatest degree of mechanization will also have the greatest energy requirement and the smallest labour input per unit of production. Thus for a given good there is a constant labour-energy product. The design of the plant alters the labour: energy ratio but not their product. Hence for highly mechanized plants γ_i is high (assuming electricity is the source of energy) while for labour intensive industries γ_i is low.

The total industrial electric energy usage in the region during interval t is given by

$$W_t^{(2)} = \sum_{i=1}^{N_t^{(2)}} \gamma_i \; \underset{\sim}{\mathrm{PROD}}_{it} \tag{5.29}$$

where $N_t^{(2)}$ = number of industrial consumers in interval t.

Because of the linear relationship of equation 5.28 equation (5.29) may be rewritten

$$W_t^{(2)} = \gamma_t \cdot \sum_{i=1}^{N_t^{(2)}} \underset{\sim}{\mathrm{PROD}}_{it} \tag{5.30}$$

where γ_t = an aggregate proportionality coefficient in interval t.

An increase in energy usage accompanies any increase in production. The way in which the increased production is achieved is important when assessing the increase in energy usage.

Increased production may be achieved simply by working existing plant for longer hours. The increase in energy requirements in interval t is then

$$\Delta' W_t^{(2)} = \gamma_t \cdot \sum_{i=1}^{N_t^{(2)}} \Delta \underline{PROD}_{it} \qquad (5.31)$$

Where $\Delta\ \underline{PROD}_{it}$ = change in production by consumer i in interval t.

There is no change in the coefficients, γ_i, because the plant has not been altered. An increase in the labour force may, or may not, be required; e.g. working overtime does not require additional labour, shift working does. In either case there is no change in the γ_i (or γ_t) or the number of consumers.

When the necessary increase in production cannot be achieved by increased utilization of plant new consumers must form. New consumers must also form if the goods demanded cannot be produced by existing consumers. Again the constraint of profitability limits the number of consumers that form in a given time interval. Growth in the demand for manufactured goods is clearly a function of population growth; hence the increase in industrial consumer numbers in interval t is

$$\Delta N_t^{(2)} = f(\Delta P_t) \qquad (5.32)$$

The form of this function is peculiar to the region. It may be modified by changes in the population's tastes and standard of living.

Each new consumer requires a labour force. If labour is scarce or, equivalently, expensive relative to the cost of energy then there is a tendency for new plant to be more highly mechanized (31). As a result the coefficients (γ_i , $i = N_t^{(2)} + 1, \ldots, N_t^{(2)} + \Delta N_t^{(2)}$) of new plants are higher than those of existing plant. Such a situation is observed when, as a result of government planning etc, the rate of growth of industrial production is increased (78). Consequently the proportional rate of increase in electric energy usage exceeds the proportional rate of increase of production. The increase in energy usage due to increased consumer numbers is given by

$$\Delta'' W_t^{(2)} = \sum_{i = N_t^{(2)} + 1}^{N_t^{(2)} + \Delta N_t^{(2)}} \gamma_i \; \underline{PROD}_{it} \qquad (5.33)$$

The total electric energy usage in interval (t + 1) is then given by

$$W_{t+1}^{(2)} = W_t^{(2)} + \gamma_t \cdot \sum_{i=1}^{N_t^{(2)}} \Delta \; \underline{PROD}_{it} + \sum_{i = N_t^{(2)} + 1}^{N_t^{(2)} + \Delta N_t^{(2)}} \gamma_i \; \underline{PROD}_{it} \qquad (5.34)$$

The above development leading to equation 5.34 emphasizes that electric energy usage is determined by the amount of production within the region. The form of the relationship is largely determined by the design of industrial plant and the availability of labour.

The coefficients, γ_i, for existing plant may be determined from historical data, e.g. using equation 5.30, while the, γ_i, for projected plants can only be determined by some form of extrapolation, unless the plant is already under construction. The likely size of the future work-force will also influence estimates of future γ_i.

Before equation 5.34 can be used to estimate future energy usage the future demand for goods, and hence the level of production in the region, must be estimated. For household perishable goods, e.g. food and clothing, the demand is a function of population size. For durable goods an analysis of households' accumulation processes, similar to that outlined in section 5.3.3 is required.

5.4.3 Evaluation of growth of non-domestic energy usage.

This section, so far, has divided non-domestic energy usage into two sectors, commercial and industrial, because of their different growth characteristics. Until recently (1969) records of energy usage have only been kept for the non-domestic sector as a whole. Consequently it is difficult to carry out any meaningful quantitative analysis of growth of energy usage.

In the preceding sections (5.4.1 and 5.4.2) the number of consumers (both commercial and industrial) has been postulated as a function of regional population size. In figure 5.4 non-domestic consumer numbers are plotted against population size for the whole of New Zealand. This plot suggests that the relationship is approximately linear, i.e.

$$N^*_t \quad = \quad h\, P_t \qquad (5.35)$$

where N^*_t = total number of non-domestic consumers in interval t,

h = constant of proportionality, and

P_t = total population in interval t.

The total value of all goods and services produced in the region in interval t is measured by the region's gross "national" product (GNP) in that interval (19). All non-domestic consumers handling a particular good contribute to its total value. Hence GNP_t is a measure of both industrial production and commercial sales volume in interval t. This suggests also that non-domestic energy usage is related to GNP. In figure 5.5 non-domestic energy usage is plotted against GNP at constant prices. This plot shows a high correlation between the variables. However, changes in GNP cannot anticipate changes in energy usage as this is a prerequisite (in some form) for production; in fact changes in energy usage would anticipate, over lead times of one or two months, changes in GNP. Consequently, if GNP is a variable in the growth model, forecasts of it must be made prior to energy usage forecasts. Ball and Burns (63) have developed an econometric model of the UK economy which exhibits good forecasting accuracy over lead times of one or two years. Over longer lead times, e.g. ten years, no alternative to trend extrapolation appears to be available.

Because it is difficult to estimate plant parameters e.g. the γ_i, and the demand for goods, over lead times of ten or more years it appears that the models, as outlined above, may be of limited usefulness. Extrapolation must in general be used to determine these parameters and unless the errors in such extrapolations were extremely low it is unlikely that forecasts of energy usage obtained by combining such extrapolations would improve upon direct extrapolations of time series data.

The models are of use in determining the effects of changes in Government economic policy, for example, on electric energy usage. A sensitivity analysis of the model is possible. This is difficult to do with time series models.

5.5 Summary

A load growth structure has been developed. The basis for the structure is the growth of consumer numbers and the average energy usage of consumers.

Increases in domestic consumer numbers are shown to be governed by population growth, subject to sufficient dwellings being available. A domestic consumer's energy usage grows as more appliances are purchased. The ordering of purchases is determined by necessity and personal taste while the rate at which appliances are purchased is governed by wealth and household population.

Non-domestic consumer numbers are determined by the size of the market for goods and hence population size. An upper limit on numbers is imposed by the requirement that each consumer's share of the market be sufficient to enable a profit to be made. A lower limit, in the case of commercial consumers, is imposed by the population's willingness to travel to purchase the desired goods. In the industrial case the availability of capital and labour also impose an upper limit on consumer numbers.

The principal determinant of commercial electric energy usage is the size of the premises which, it is shown, is a function of the size of the work-force employed. Industrial energy usage is a function of the amount of production. Plant design determines the amount of electric energy required for each unit of production. Non-domestic consumers do not accumulate appliances in the way domestic consumers do. Increases in an established non-domestic consumer's energy usage occur as a result of increased usage of existing appliances, i.e. by increasing their appliance usage factors.

In the next chapter the short term behaviour of the load is discussed. The determinants of appliance usage patterns and hence appliance usage factors are examined in some detail.

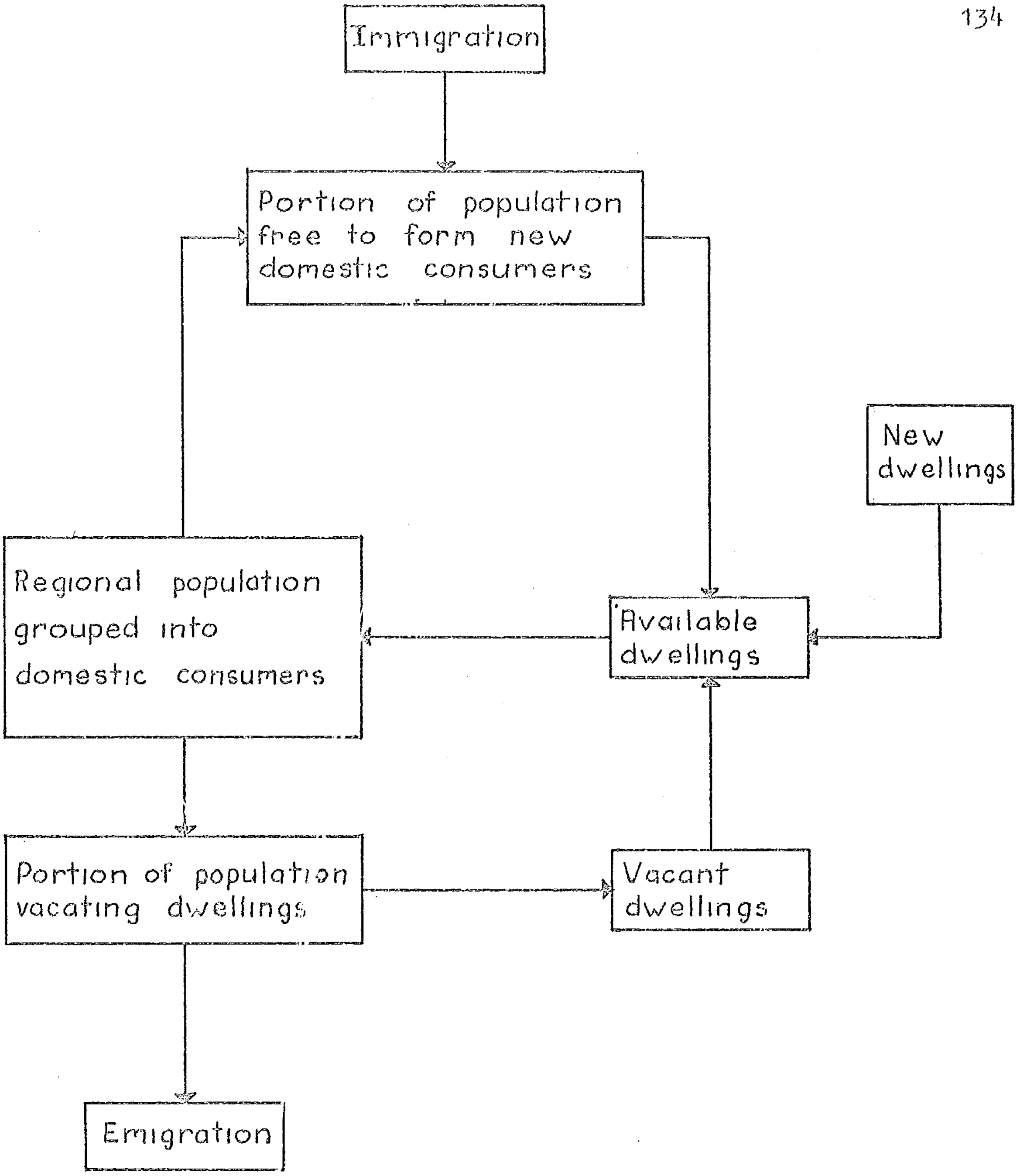


Figure 5.1 The growth of domestic consumer numbers.

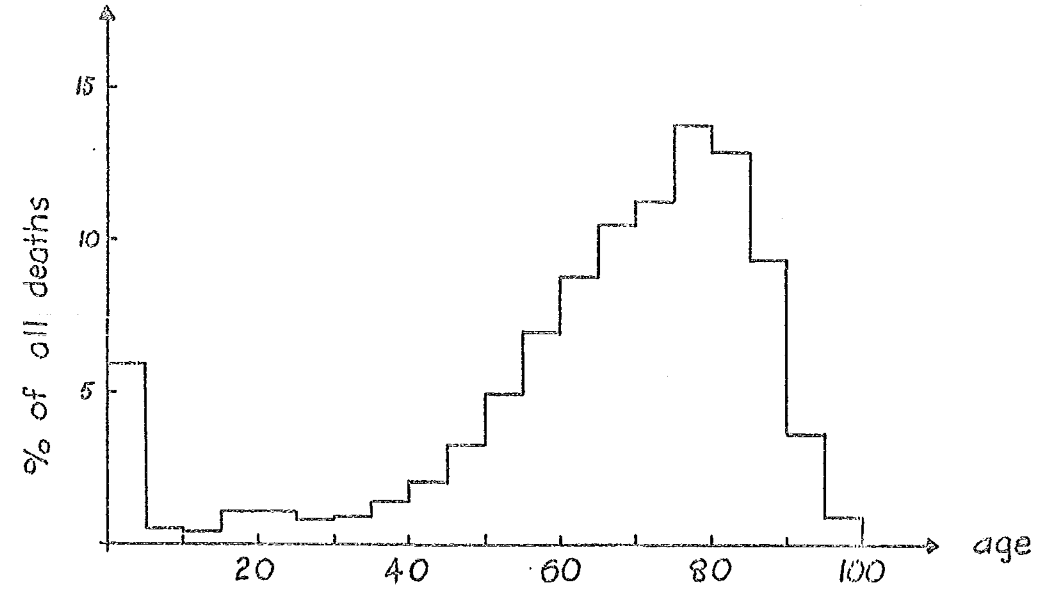


Figure 5.2(a) Age group at death (in 1967)

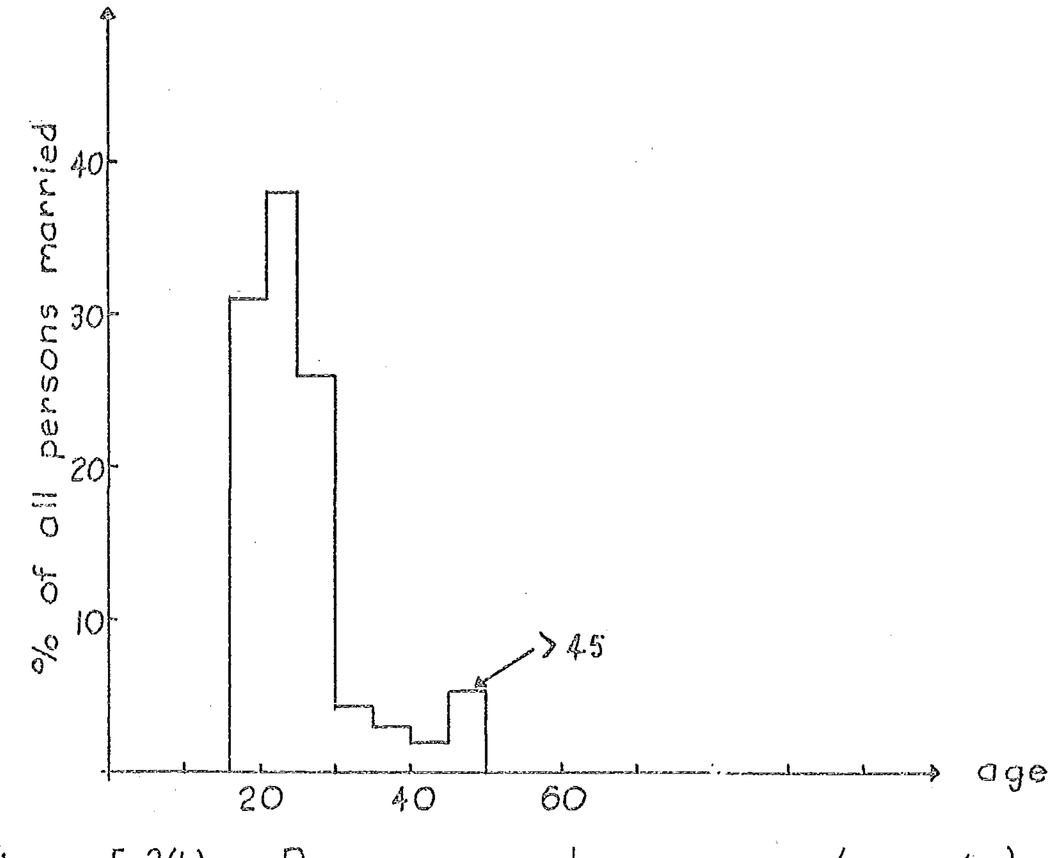


Figure 5.2(b) Age group at marriage (in 1967)

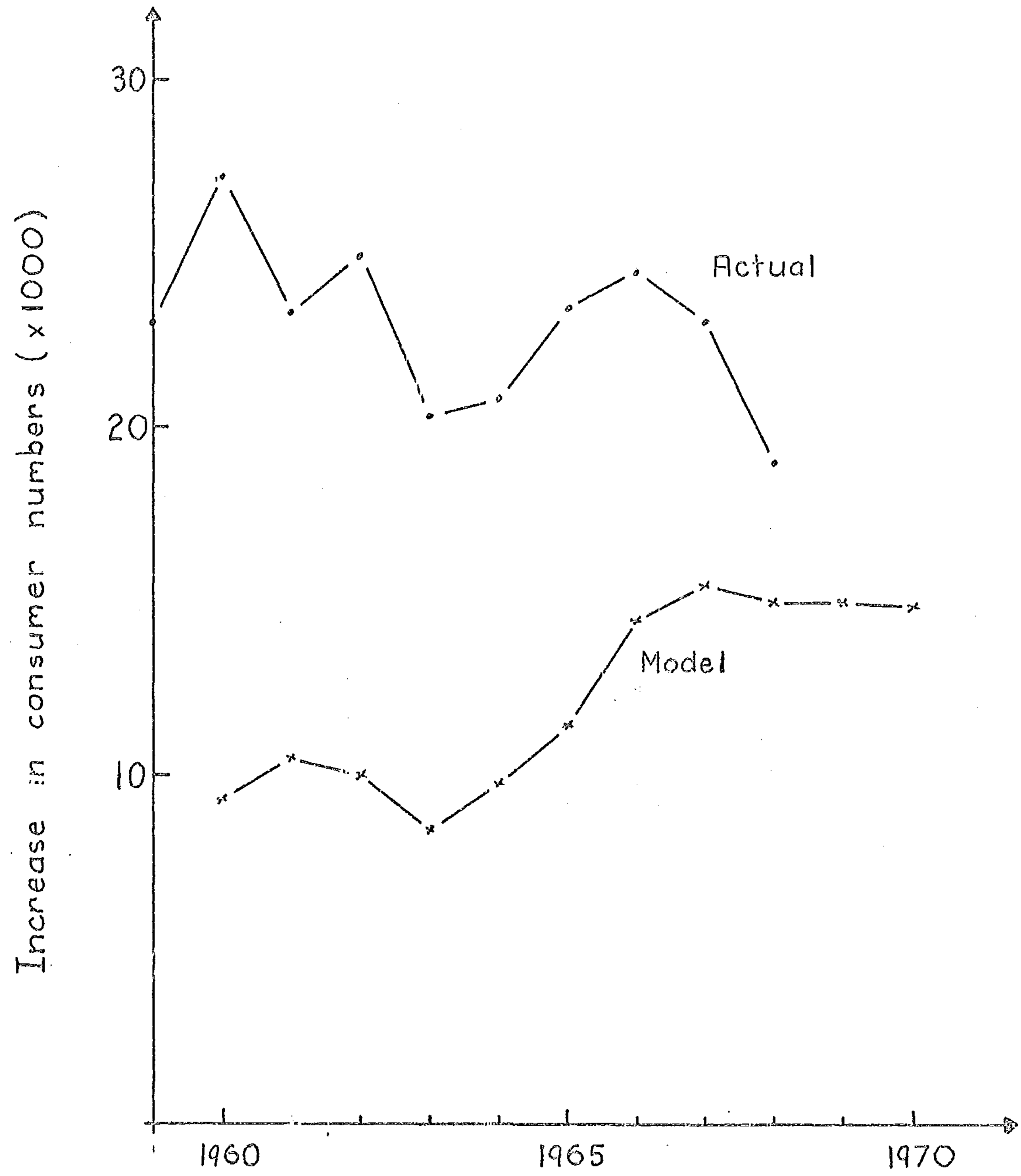


Figure 5.3 Growth of domestic consumer numbers : comparison between "birth and death" model and actual occurrences.

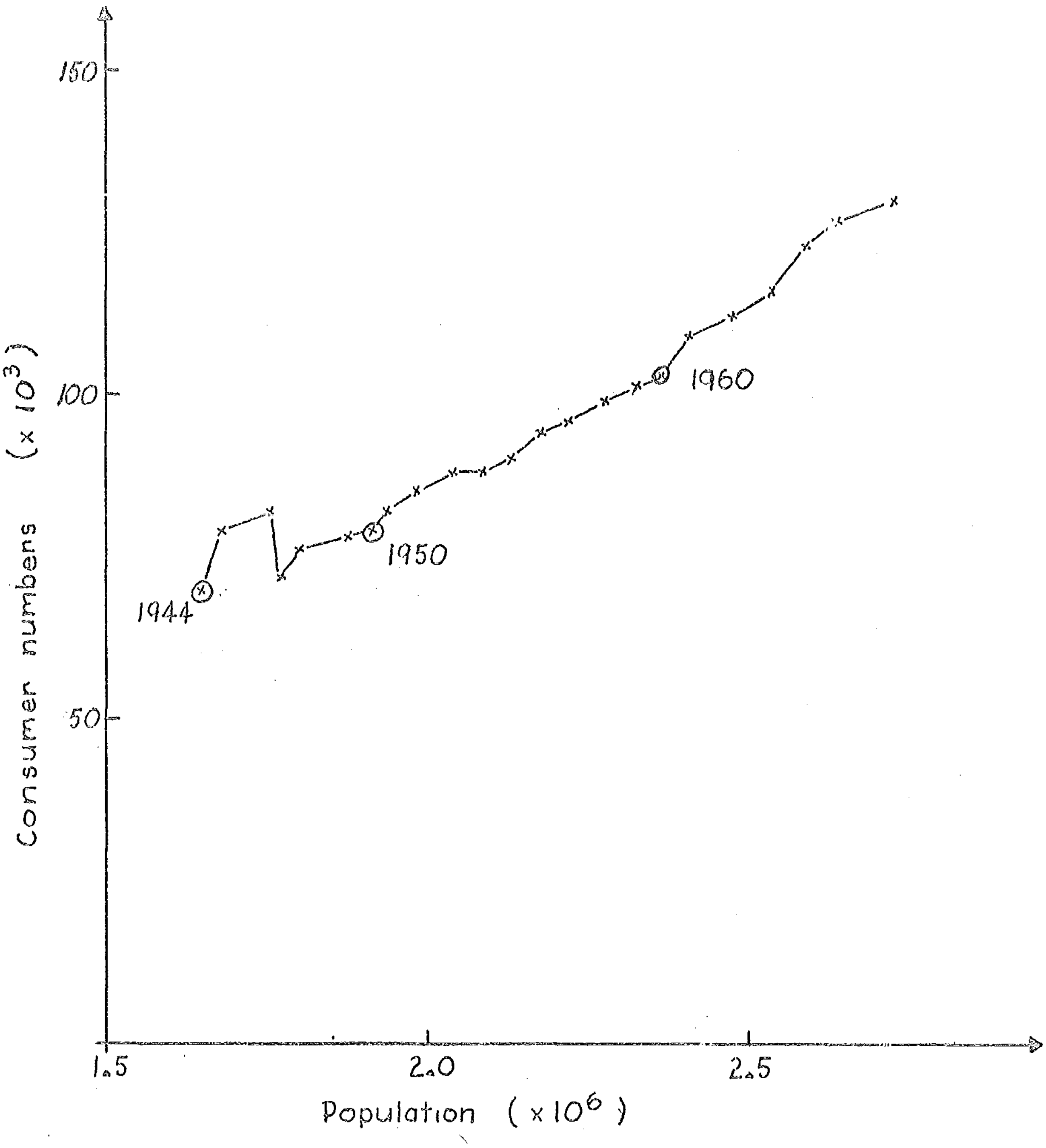


Figure 5.4 Non-domestic consumer numbers as a function of population in New Zealand

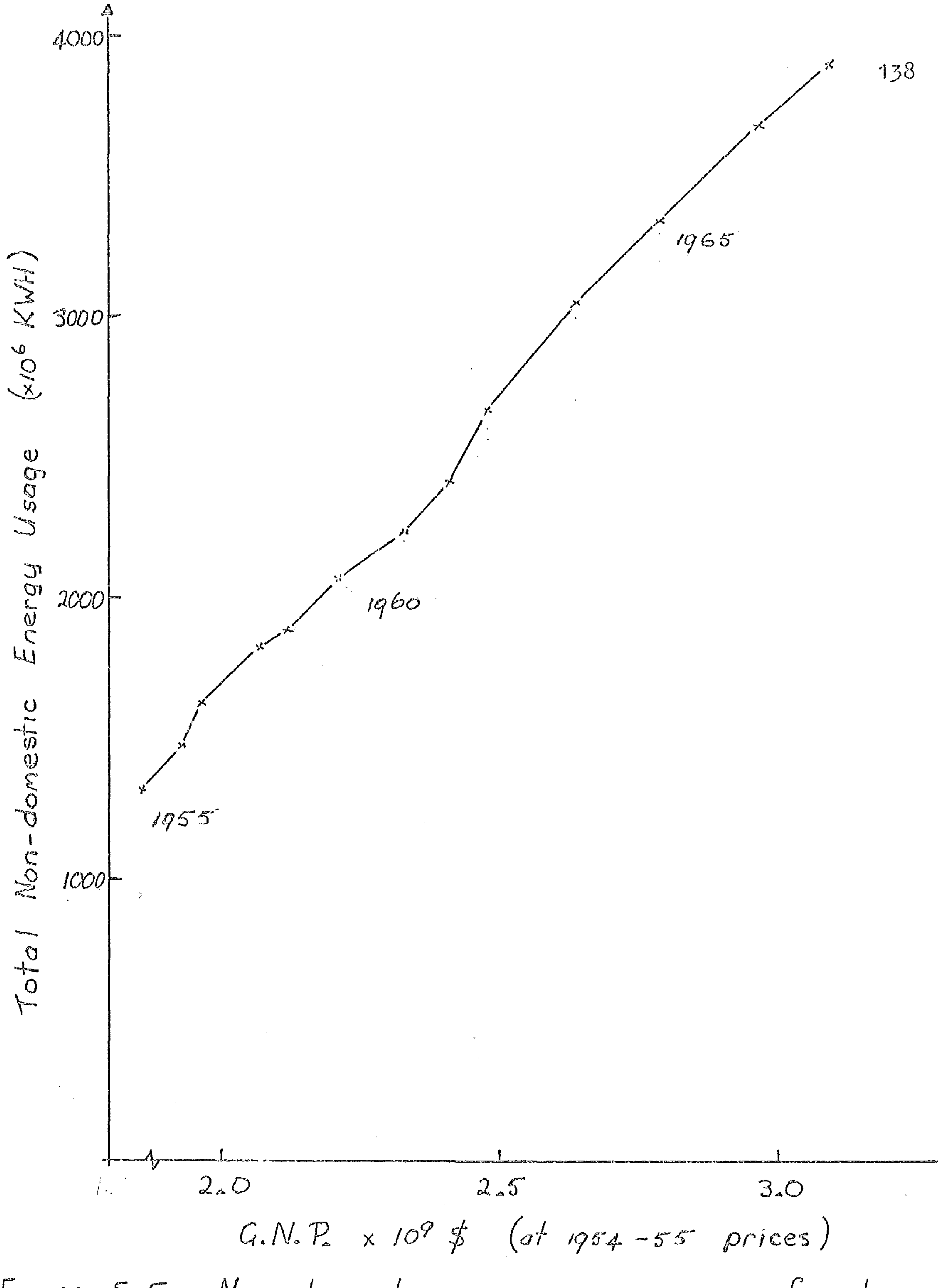


Figure 5.5 Non-domestic energy usage as a function of Gross National Product.

CHAPTER 6

SHORT TERM LOAD BEHAVIOUR

6.1 Introduction

The short term behaviour of the load is determined by the way consumers use their appliances. In this context "short term" means from a few hours to a year. It was shown in Chapter 2 that the shape of the daily load curve is determined by the daily sequences of probabilities that appliances of each type are connected at each interval of the day. The current chapter is primarily concerned with the determination of these probabilities from the available data. The daily sequence of probabilities multiplied by the demand factor for each type of appliance forms a load curve component. Thus the determination of the probabilities is equivalent to identifying and separating the load curve components from the complete load curve.

A single component can only be identified if its contribution to the daily load curve varies relative in some sense to the other components (79). This means that if several components vary with the same independent variable, e.g. time of day, in the same relative proportions, they cannot be separated. Hence identification of the load components requires that appropriate transformations be applied to the data. In some circumstances the load curve will not separate into one component for each appliance type; e.g. the usage of all industrial appliances is determined by a common independent variable (time of day) and while their total contribution to the daily load curve can be separated from that of other appliances individual appliance contributions cannot. This is not a major restriction; it merely means that the components which are separated are aggregates of other components.

Three models of the short term load behaviour are developed in this chapter. In the first the seasonal behaviour of the daily load curve is synthesized from three artificial component load curves. The model represents the load behaviour in average environmental conditions. Load curves for entire seasonal cycles can be generated with this model; it forms a useful tool for power system planning studies.

A second model is based on Fourier series representation of the integrated demand in each interval of the day. It also represents load behaviour in an average environment. Generation of long sequences of load curves is also possible.

The third model is based on the relationship between demand and the climatic variables, in particular temperature and hours of daylight. Using this model the load behaviour can be reproduced for any given temperature conditions and hours of daylight. Forecasting over lead times of a few hours to one or two days is the major application. It is shown that unless weather forecasting accuracy is improved and, or, the model made more accurate it is not possible to achieve the accuracy desirable in short term demand forecasts.

6.2 Modelling the seasonal behaviour of the load

6.2.1 A physical explanation of seasonal behaviour

Seasonal changes in the way consumers use their appliances are manifest as changes in the shape and magnitude of the daily load curve. These changes have been illustrated earlier, in figure 2.2, for an urban supply authority load of about 71000 domestic and 9000 non-domestic consumers. The seasonal variation in daily energy usage for this supply authority is illustrated in figure 6.1; this variation is approximately sinusoidal with a period of one year.

The load curves of figure 2.2 exhibit two peaks which occur at the same time each day, i.e. 0830 hours and 1730 hours. In mid-summer a third, subsidiary, peak occurs at about 2030 hours. It occurs progressively earlier as winter approaches until it merges with the peak at 1730 hours. The occurrence of this subsidiary peak coincides roughly with the time of sunset. It is postulated that this peak is caused by lighting. Lighting is turned on around the time of sunset and remains in use until some fairly regular time, e.g. 2200-2300 hours. As the sun sets earlier lights are turned on earlier but are still turned off at about the same time. The amount of lighting brought into use is determined by the numbers of rooms in use in a dwelling. Hence the lighting load will tend to be larger in early evening, when more rooms are in use.

Lighting is also used if it is not fully daylight at the time when the population gets up in the morning, i.e. about 0630-0700 hours on weekdays. This has the effect of increasing the rate of rise of demand in the morning.

Heating appliances are brought into use when the internal temperature of buildings falls below some "comfort-level" (80,81), which can vary between individuals. In the majority of households heating appliances are not used when the occupants are in bed, regardless of temperature. The heating load first appears in the seasonal cycle in late evening as temperatures drop. As winter approaches it appears when the population gets up, remains on until the day is sufficiently warm and then reappears as temperatures drop toward sunset and then remains on until the population goes to bed. Again the amount of heating in use in a dwelling at any particular time depends on the number of rooms in use.

6.2.2 A three component model of the normalized daily load curve.

In this section a model which synthesizes the seasonal changes in load curve shape from three separate components is developed. To separate shape changes from magnitude changes the daily load curves may be normalized with respect to daily energy usage. If D_{jt} denotes the integrated demand in the t th interval on day j then the corresponding normalized demand D^*_{jt} is given by

$$D^*_{jt} = (D_{jt} / \sum_{t=1}^{M} D_{jt}) \cdot L \qquad (6.1)$$

where M = number of daily intervals, e.g. 48 half hourly intervals

L = any constant, usually 1 or 100

A three component model was postulated to explain the observed seasonal changes in the shape of the normalized daily load curve. The three components are:

1. A base component representing the constant shape portion of the load curve; this component has two seasonally stationary peaks at 0900 and 1730 hours.
2. An additive morning component with a peak at 0700 hours and a magnitude that varies sinusoidally with the time of the year.
3. An additive evening component of constant magnitude but varying duration. The duration is assumed to vary sinusoidally with the time of the year.

Mathematically this model of the normalized integrated demand in internal t on day j is written;

$$\hat{D}^{*}_{jt} = a_1 f_{1t} + a_2 \left(1 + \sin\left(\frac{2\pi j}{N} + \omega_2\right)\right) f_{2t} + a_3 f_{3t}(j) \qquad (6.2)$$

where N = number of days in the seasonal cycle (assumed to be 365),

f_{1t} = base component of constant shape,

f_{2t} = morning component also of constant shape,

$f_{3t}(j)$ = evening component with a shape dependent on j ,

a_1, a_2, a_3 = scale factors, $a_1, a_2, a_3 \geqslant 0$, and

ω_2 = phase shift constant to align the model with the seasonal cycle.

In the model each component is represented as a tabulated function of the daily interval t ; in figure 6.2 the basic shapes of the components are plotted. The use of tabulated functions is particularly useful when the model is implemented on a digital computer as the shape of the components can be easily changed to represent a different load.

The evening component $f_{3t}(j)$ has the following properties. For intervals $t \geqslant t_1$, see figure 6.2, the shape remains constant with j ; i.e.

$$f_{3t}(j) = f_{3t}(1) \qquad , t_1 \leqslant t \leqslant M \qquad 6.3)$$

Interval t_1 is the interval immediately following the latest time of sunset. The time of sunset varies with j ; the variation is approximately sinusoidal with a period of one year. The shape of the function $f_{3t}(1)$ in figure 6.2 for $t < t_1$ represents the way in which lighting is brought into use. This portion of $f_{3t}(1)$ is shifted left on the time axis by an integral number of intervals, τ ($\tau \geqslant 0$), where τ is the integer obtained by truncating the quantity

$$a_4 \left(1 + \sin\left(\frac{2\pi j}{N} + \omega_4\right)\right) \qquad (6.4)$$

where a_4 = a scaling factor , $a_4 \geqslant 0$, and

ω_4 = phase shift constant to align the model with the seasonal cycle.

Hence

$$f_{3t}(j) = f_{3(t+\tau)}(1) \qquad 0 < t \leqslant (t_1 - \tau) \qquad (6.5)$$

In the time between sunset (at $t_1 - \tau$) and time t_1

$$f_{3t}(j) = f_{3t_1}(1) \quad , \quad (t_1 - \tau) < t < t_1 \qquad (6.6)$$

Typical variations of $f_{3t}(j)$ with j are shown in figure 6.3.

Clearly whether the model successfully represents changes in load curve shape rests on the specification of the component shapes and the associated scale factors a_1, a_2, a_3, a_4. The component shapes may be determined from historical data by a mathematical technique, if one is available, or from experience in a way analogous to that of Reps (67). Given the components, the scale factors may be determined by fitting the model to actual data.

6.2.3 Experimental verification of the three component model.

The model of seasonal load behaviour represented by equations 6.2 to 6.6 was applied to the new data shown in figure 6.4. This data is the set of consecutive Tuesday load curves in the year 27.3.67 - 31.3.68 for the urban load of figure 2.2 By using only Tuesday load curves day-of-week effects are avoided. Public holidays occurred on the 5th and 40th Tuesdays, and the 41st occurred during the annual summer holidays; on these days the industrial load was not contributing to the demand. The weather on the 28th Tuesday was unusual for the time of the year and the load curve on this day had an unusual shape. Consequently these four days were ignored in the experiment.

The remaining 49 daily load curves were normalized such that

$$\sum_{t=1}^{M=48} D^{*}_{jt} = 100.0 \quad \text{for } j = 1, \ldots, 49 \qquad (6.7)$$

For each interval t of the day the 49 observations of the normalized demand were averaged, a standard deviation computed and the maximum and minimum values of normalized demand determined. These four quantities are plotted for each daily interval in figure 6.5.

The shapes of the model components in figure 6.2 were determined from the mean curve of figure 6.5, also plotted in figure 6.2 , as follows;

1. The mean curve was modified over the intervals 0600 to 0900 to shift the morning peak to 0900 hours.
2. The subsidiary evening peak on the mean curve was removed to give a constant drop in demand for each interval from 1900 to 2400 hours.

The curve remaining after steps 1. and 2., when tabulated, forms the base component.

3. The base component was subtracted from the mean curve.

The morning residual from step 3 formed the f_{2t} component in figure 6.2 while the evening residual formed the $f_{3t}(1)$ component.

The scaling factors were found by minimizing the functional

$$F = \sum_{j=1}^{N} \sum_{t=1}^{M} \left| D^*_{jt} - \hat{D}^*_{jt} \right| . \qquad (6.8)$$

Where N = length of seasonal cycle in days, and

M = number of intervals in the day.

The unconstrained minimization technique due to Rosenbrock (82) was used to perform the actual minimization.

The normalized daily load curves generated by the model (equation 6.2) were then subtracted from the normalized raw data to give a set of daily error curves; i.e.

$$e_{jt} = D^{*}_{jt} - \hat{D}^{*}_{jt} \qquad (6.9)$$

$$t = 1, \ldots, M = 48$$

$$j = 1, \ldots, N = 49$$

Note that

$$\min (F) = \sum_{j=1}^{49} \sum_{t=1}^{48} |e_{jt}| \qquad (6.10)$$

At each interval of the day the maximum and minimum values of e_{jt} were determined and the mean computed. If the model represented the data perfectly all three values would be zero for all intervals. In fact the model is not perfect; it was considered satisfactory if the mean was zero at each interval of the day.

The initial choice of component shapes, as in figure 6.2, did not give zero mean error in all intervals of the day. The shapes were modified manually, to reduce the mean error, and new scaling factors computed together with a new set of daily error curves. After five repetitions of this process no further improvement appeared possible. In figure 6.6 the final shapes of the components and the final scale factors are shown. In figure 6.7 the mean, maximum and minimum values of the error expressed as percentages of the daily energy usage are plotted.

These experimental results show that changes in the shape of the daily load curve can be represented using only three components with reasonable accuracy. However experimentation is needed to determine the shapes of the model components to give the best agreement with actual data. The existing model only represents changes in daily load curve shape. It contains no information as to the magnitude of the demand. The seasonal changes in daily load curve shape must be regular, e.g. sinusoidal. Variations which are a result, for example, of abrupt weather changes cannot be represented.

6.2.4 An extension of the three component model.

The model described above is based on load curves normalized with respect to daily energy usage. The model can be restored to units of demand simply by scaling each curve by the appropriate daily energy usage value. A suitable model of the daily energy usage on day j of the year is suggested by figure 6.1 to be

$$W_j = W_b + \frac{W_s}{2} \sin\left(\frac{2\pi j}{N} + \omega\right) \quad j = 1, \ldots, N \qquad (6.11)$$

Where W_b = annual average daily energy usage,

W_s = peak to peak seasonal variation in daily energy usage, and

ω = phase shift constant to align the model with the seasonal cycle.

The resulting model of the integrated demand in interval t on day j becomes

$$\hat{D}_{jt} = \hat{W}_j \cdot \hat{D}^*_{jt} \tag{6.12}$$

Estimates of the parameters W_b and W_s may be determined from historical energy usage data as follows;

$$\hat{W}_b = \frac{1}{N} \sum_{j=1}^{N} W_j \tag{6.13}$$

$$\hat{W}_s = \frac{4}{N} \sum_{j=1}^{N} W_j \sin \left(\frac{2\pi j}{N} + \omega \right)$$

Note that $\frac{W_s}{2}$ is the coefficient of the first sine term in the Fourier expansion of the daily energy usage data.

In figure 6.8 the sequence of daily load curves generated by the model of equation 6.12, with parameters calculated for the data of figure 6.4, is plotted. Figure 6.8 may be compared with the actual load curves of figure 6.4. The model tends to over-emphasize the magnitude differences and underemphasize changes in shape. That the model generates a smoothed sequence of load curves can be seen by comparing figure 6.8 with the mean weekday curves of figure 2.2.

6.3 Determination of the components of the daily load curve

Each component of the model described above represents a contribution from a distinct group of appliances. The shape of each component is a measure of the probability of finding appliances in that group connected in each interval of the day.

These probabilities were determined on a trial and error basis in the above model. In this section two mathematical techniques for identifying components directly from the raw data are discussed.

6.3.1 Component identification using the Karhunen-Loeve Expansion.

The Karhunen-Loeve expansion, which is discussed in chapter 4 and Appendices B and C, resolves an ensemble of sample functions into a set of orthogonal functions with uncorrelated coefficients. Using the discrete form of the expansion the integrated demand in interval t on day j may be represented exactly (83) by,

$$D_{jt} = \sum_{i=1}^{M} a_{ji} \lambda_i^{\frac{1}{2}} \phi_{it} \qquad (6.14)$$

Where a_{ji} = uncorrelated coefficient

The λ_i, $\{\phi_{it}, t = 1, \ldots, M\}$ $i=1, \ldots, M$ are the eigenvalues and eigenvectors respectively of the covariance matrix

$$\{R_{t,s}\} = \underset{j}{E} \{ D_{jt} \; D_{js} \} \qquad t, s = 1 \ldots, M \qquad (6.15)$$

The spectral coefficients, a_{ji}, are obtained from

$$a_{ji} = \frac{1}{\lambda_i^{\frac{1}{2}}} \sum_{t=1}^{M} D_{jt} \phi_{it} \qquad (6.16)$$

for $j = 1, \ldots, N$

$i = 1, \ldots, M$

The product $(a_{ji} \cdot \lambda_i^{\frac{1}{2}})$ is a measure of the energy associated with the component ϕ_i on day (72).

It was hypothesized that the orthogonal functions for an ensemble of load curves over an entire seasonal cycle would be the desired component load curves. If this were so the associated a_{ji} would behave in some consistent fashion for $j = 1, \ldots, N$, representing the seasonally changing contribution of that component.

For the ensemble of load curves in figure 6.4 the $\left\{ \phi_{it},\ t=1, \ldots, M \right\}$, λ_i and a_{ji} for $i = 1, \ldots, M$, $j = 1, \ldots, N$ were calculated. The ϕ_{it} and a_{ji} corresponding to the three largest λ_i are plotted in figures 6.9 (a) and (b) respectively. Note that ϕ_{1t} has the shape of the average daily load curve for the ensemble while the a_{1j} varies with j in a similar way to the daily energy usage in figure 6.1. However, the a_{ji} for $i > 1$ in figure 6.9 (b) do not exhibit behaviour consistent with the hypothesis that the associated ϕ_{it} represent seasonally varying component loads.

Further consideration of the nature of the Karhunen Loeve expansion shows that it can only distinguish load curve components in certain cases. As a simple illustration consider the case where there are only two integrated demand intervals in each day. Each daily load sequence may then be treated as a vector in two dimensional space; the vector for day j is denoted by $\underset{\sim}{V}_j$. Three such vectors are drawn in figure 6.10(a). The directions of each vector are determined by the shape of the load curve. The vectors of figure 6.10(a) resolve into orthogonal vectors;

$$\begin{aligned} \underset{\sim}{\psi}_1 &= \lambda_1^{\frac{1}{2}} \quad \underset{\sim}{\phi}_1 \\ \underset{\sim}{\psi}_2 &= \lambda_2^{\frac{1}{2}} \quad \underset{\sim}{\phi}_2 \end{aligned} \tag{6.17}$$

Vector $\underset{\sim}{\psi}_1$ is a measure of the common components of the load vectors, $\underset{\sim}{\psi}_2$ a measure of the vector differences. In figure 6.10(b) the load vectors have similar directions, reflecting similar load curve shapes. The magnitude of $\underset{\sim}{\psi}_2$ is considerably reduced.

These two simple examples illustrate the point that the Karhunen Loeve expansion can identify load components only if the components alter the shape of the load curve. As the statistical summary of figure 6.5 shows, the change in shape of the daily load curve is relatively small. Any seasonal changes that do occur are swamped by variations arising from changes due to weather etc. Hence the apparently random variation of the a_{ji} in figure 6.9(b).

6.3.2 Component identification using Fourier analysis.

A plot of the integrated demand at a particular interval of the day for consecutive days of the year exhibits an approximately sinusoidal variation, e.g. curve (a) in figure 6.12 which is drawn for interval 40, i.e. 2000 hours. This suggests that a Fourier analysis of the data may identify component loads as higher frequency terms. The Fourier series model of the demand at interval t on day j is

$$\hat{D}_{jt} = a_{ot} + \sum_{i=1}^{K} a_{it} \cdot \sin\left(\frac{2\pi i}{N} \cdot j\right) + \sum_{i=1}^{K} b_{it} \cos\left(\frac{2\pi i}{N} j\right) \tag{6.17}$$

Where N = length of seasonal cycle in days.

Estimates of the coefficients in this model can be obtained from raw data as follows;

$$a_{ot} = \frac{1}{N} \sum_{j=1}^{N} D_{jt}$$

$$a_{it} = \frac{2}{N} \sum_{j=1}^{N} D_{jt} \sin \left(\frac{2\pi i}{N} j\right) \qquad (6.18)$$

$$b_{it} = \frac{2}{N} \sum_{j=1}^{N} D_{jt} \cos \left(\frac{2\pi i}{N} j\right)$$

$$\text{for } i = 1, \ldots, K$$
$$t = 1, \ldots, M$$

These coefficients were calculated for each daily interval over N = 371 consecutive days with K = 100; the load was the same as that in figure 2.2. These calculations revealed that the sine coefficients for i = 1 and i = 53 were the most significant at all intervals of the day. The sequence a_{1t} for $t = 1, \ldots, M$ represents the additive seasonal component while the a_{53t} component represents a contribution which varies with the day of the week. The sequences a_{0t}, a_{1t} and a_{53t} for $t = 1, \ldots, M = 48$ are plotted in figure 6.12.

To illustrate the mismatch between the raw data and Fourier series model daily error curves for the data of figure 6.4 were calculated by

$$e_{jt} = D_{jt} - \left(\hat{a}_{0t} + \hat{a}_{1t} \sin \frac{2\pi i}{N}\right) \quad \begin{array}{l} j=1,\ldots, N \\ t=1,\ldots, M \end{array} \qquad (6.19)$$

Where $N = 53$, the number of days in the seasonal cycle of Tuesdays, and

$\hat{a}_{0t}$, $\hat{a}_{1t}$ were the sequences plotted in figure 6.11.

The sequence of e_{jt} for $j = 1, \ldots, 53$ in interval 36 ($t = 36$) is plotted in curve (b) in figure 6.11. If the model were perfect curve (b) would be zero for all j. The deviations from zero may be attributed to

(a) deviations of the weather from the seasonal norm, and

(b) holidays.

Thus the model of the demand in equation 6.19 is a smoothed model. For the purposes of comparison with the three component model a sequence of 53 daily load curves was generated by

$$\hat{D}_{jt} = \hat{a}_{0t} + \hat{a}_{1t} \sin \frac{2\pi j}{N} \qquad (6.20)$$

$$j = 1, \ldots, N = 53$$

$$t = 1, \ldots, M = 48$$

These load curves are plotted in figure 6.13. This figure may be compared with figure 6.8. It is apparent that the Fourier series model generates a much smoother set of load curves. Both model generated figures may be compared with figure 2.2(b). This comparison shows that both models represent the average behaviour of the daily load curves.

Models such as those described so far treat the seasonal demand variation as being sinusoidally dependent on time alone.

Provided the determinants, e.g. temperature, of the seasonal variation also vary sinusoidally, acceptable agreement between model and data is obtained. When the determinant variables deviate from their assumed sinusoidal behaviour the models break down, as the various figures show. An analysis of the dependence of demand on temperature and daylight illumination is the subject of the next section.

6.4 Climate and short term load behaviour

6.4.1 Temperature sensitivity of the load

The form of the temperature sensitivity of the load is illustrated in the scatter diagrams of Appendix D. In these diagrams the integrated demand data of figure 6.4 (for the second half hour of each hourly period) is plotted against the ambient air temperature measured on the hour.

The main feature of these plots is a well defined breakpoint which occurs about a temperature of 60°F for most of the day. This breakpoint occurs at a slightly lower temperature (about 55°F) during the hours 0100 to 0600. The temperature at which these breakpoints occur corresponds to the "comfort thresholds" determined by Bastings and Simmons (81) of 60°F for living rooms in dwellings and about 54°F for bedrooms. The comfort threshold varies between different countries; c.f. 65°F in Britain and 72°F in the U.S.A. (80, 84).

Demand rises approximately linearly with decreasing temperature below the breakpoint, while above it there is relatively little change in demand. Most of the very low observations in Appendix D occurred during the Christmas holiday period.

6.4.2 Demand - temperature models

The temperature sensitive portion of the load is composed mainly of heating appliances. These are used, in general, to modify the internal temperature of buildings. Hence the demand is sensitive to the internal temperature of buildings rather than the ambient air temperature. Denote the internal temperature during interval t, on day j, by Θ_{jt}. The relationship between Θ_{jt} and the corresponding external temperature, denoted by T_{jt}, is determined by the construction of the building, i.e. materials and amount of ventilation, and its orientation with respect to sunshine (14, 84, 85). In general the internal temperature lags changes of external temperature; the amount of lag depends on the building's construction and may vary widely between buildings (84). Consequently the assessment of the time lag appropriate to a particular load, and its mixture of building types, must be based primarily on a data analysis.

A simple, empirical, expression for the internal temperature assumes the rate of change of internal temperature proportional to the temperature difference across the walls of the building; i.e.

$$\frac{d\Theta}{d\tau} = \eta (T - \Theta) \qquad (6.21)$$

Where T = ambient (external) air temperature,

Θ = internal temperature, and

η = reciprocal of the thermal time constant.

The solution of this equation in continuous form is (see Davies (14)),

$$\Theta = \eta \int_0^{\infty} \exp(-\eta \tau) \; T(\tau) \; d\tau \qquad (6.22)$$

Where τ = 0 is the present time and is measured toward the past, and

Θ = internal temperature at time τ = 0

After replacing the integral by summation over sufficiently small time intervals and making the substitution

$$\rho = \exp(-\eta) \qquad (6.23)$$

the expression for internal temperature at time t , Θ_t, becomes

$$\Theta_t = (1-\eta) \sum_{i=0}^{K} \rho^i T_{t-i} \qquad (6.24)$$

Where K = a constant such that the ambient temperatures in intervals further than K intervals into the past have negligible effect on the internal temperature at time t .

A correlation analysis performed by Davies (14) to determine ρ gave a value $\rho = 0.955$, which corresponds to a thermal time constant of about 22 hours; i.e, K = 22. These values apply to English conditions and are not necessarily appropriate in New Zealand because of differences in building construction. However Bastings (84) compares the thermal transmittance (U), which is defined as the rate of heat transfer from the internal air to external air (B.t.u. per square foot per hour per ^{o}F), of buildings in England and New Zealand. In England, typical values of U are 0.2, in New Zealand $0.4 < U < 0.6$, i.e. two to three times greater. Hence the thermal time constant for New Zealand loads is likely to be about half of the figure obtained by Davies; i.e. about 10 hours. In the absence of better information a value of $\rho = 0.9$ was adopted for preliminary studies.

Some evidence of the true value of the thermal time constant in New Zealand is contained in Appendix D. On the 28th Tuesday of the year, i.e. curve 28 in figure 6.4, high morning temperatures occurred; about 10 a.m. temperatures dropped suddenly; see curve (a) of figure 6.14. After some time lag the demand began to rise as in curve (b) of figure 6.14. In Appendix D the demand-temperature pairs on day 28 are identified for the hours 0800 to 1700 hours. An estimate of the time the demand took to reach the level appropriate to the lower temperature can be obtained by counting the hours from the time of the temperature drop until the point merged with other points; this is about 8 hours. This result confirms the choice of $\rho = 0.9$ as a reasonable value to use.

In figure 6.15 a scatter diagram of demand against the internal temperature of buildings (calculated from equation 6.24 with K = 21 hours) is shown for interval 40, i.e. 1930 - 2000 hours. This diagram, like those in Appendix D, exhibits a break-point at about $60^{\circ}F$, with the demand varying linearly with temperature above and below this point. With this information the following model of the demand-temperature relationship was proposed:

$$\begin{aligned} \hat{D}_{jt} &= a_{Bt} + a_{\theta t}(\Theta_{jt} - 60), & \Theta_{jt} &< 60^{\circ}F \\ &= a_{Bt} & \Theta_{jt} &\geq 60^{\circ}F \end{aligned} \qquad (6.25)$$

Where θ_{jt} = internal temperature in interval t on day j, obtained from equation 6.24 with $\rho = 0.9$, and

a_{Bt}, $a_{\theta t}$ = constants to be determined.

The parameter a_{Bt} represents the constant base load which exists for all temperatures, $a_{\theta t}$ is a measure of the sensitivity of the load to temperature changes. Both parameters must be estimated from actual data by, for example, regression analysis.

A regression analysis was performed on the demands from figure 6.4 and the associated internal temperatures. Those data points which occurred on public holidays and during the annual summer holidays, e.g. those denoted by a square in figure 6.15, were ignored in the analysis. In figure 6.16 the estimates of the coefficients obtained are given.

Using this model of equation 6.25 much of the seasonal variation in demand is explained. This is shown in figure 6.17(b) where curve (i) is the variation of integrated demand in interval 40 about the annual mean for that interval. Demands for the days ignored in the regression analysis are set to zero. Curve (ii) shows the result of correcting of temperature sensitive load; in the ideal case this curve should be a straight line. Because the model does not account for all the variation curve (ii) exhibits some residual fluctuation. For comparison figure 6.17(a) shows the same curves but instead of calculating the internal temperature from the lagged series it was assumed equal to the external temperature, i.e.

$$\theta_{jt} = T_{jt} \qquad (6.26)$$

The comparison between figures 6.17(a) and 6.17(b) highlights the influence temperature history has on consumers' decisions to use heating appliances.

The present model does not account for changes in lighting requirements depending on whether it is daylight or dark. In the next section the model is extended to include this component of the load.

6.4.3 A demand-temperature-daylight model

Some demand from lighting is always present from sunset on one day to sunrise on the next. Whether it is present during daylight depends on the amount of cloud cover, the sun's elevation, e.g. see Bastings (86) or Davies (14), and in some cases the design and construction of buildings (e.g. whether they have windows etc). Seasonal variation in the demand from lighting in a particular daily interval generally only occurs if it is possible for the sun to be either set or risen in that interval for some portion of the seasonal cycle. Consequently seasonal changes are restricted to the daily intervals between and including the earliest and latest time of sunrise and the earliest and latest time of sunset.

The contribution of lighting to the demand can be assessed manually from the scatter diagrams, e.g. figure 6.15, for the relevant daily intervals. The demand-internal temperature plots which occurred while interval 40 was in daylight are indicated by circles in figure 6.15. The points in figure 6.15 now appear as two distinct groups, both with similar behaviour (i.e. a breakpoint at $60^{o}F$ and otherwise similar slopes) but separated by about 10 MW. Assuming this is due to lighting it corresponds to each of the 70,000 domestic consumers using about 100 watts of lighting at this time, which is a reasonable figure.

Denote the variable lighting component in interval t on day j by I_{jt}. For those daily intervals when the sun is always set or risen, regardless of the time of the year, no seasonal variation due to lighting can be observed and $I_{jt} = 0$. In the remaining intervals $I_{jt} > 0$ if the sun is set, $I_{jt} = 0$ if it is risen. The demand-temperature-daylight model may now be written.

$$
\begin{aligned}
\hat{D}_{jt} &= a_{Bt} + a_{\theta t} (\theta_{jt} - 60) + I_{jt}, & \theta_{jt} < 60^{o}F \\
&= a_{Bt} + I_{jt} & \theta_{jt} \geqslant 60^{o}F
\end{aligned}
\qquad (6.27)
$$

$$
\text{for } t = 1, \ldots, M \qquad j = 1, \ldots, N
$$

Estimates of the model parameters a_{Bt}, $a_{\theta t}$ must again be obtained from the data; the component I_{jt} may be estimated from scatter diagrams as described above. Alternatively regression analysis using dummy (0, 1) variables may be used (88, 89); it was considered that there was insufficient data in this case to obtain reliable estimates of the dummy regression coefficients.

The seasonally varying lighting component must be compensated for before a regression analysis to find $a_{B}t$, $a_{\theta t}$ is performed. If it is not compensated for the parameter $a_{\theta t}$ tends to be increased due to the greater range of demands; this can be seen from an examination of figure 6.15.

Accordingly a lighting demand of 10 MW, obtained manually from scatter diagrams, was added to the raw demands in the daily intervals 13-15 and 33-40 on those days when the sun was risen during these intervals. Although the sun rises before interval 13 (0630 hours) for part of the year there is little seasonally varying lighting demand before this time. The estimates of the parameters a_{Bt}, $a_{\Theta t}$ obtained from a regression analysis performed on this compensated data are plotted in figure 6.18. These may be compared with those in figure 6.16. As expected the parameter $a_{\Theta t}$ tends to be smaller in the intervals where compensation for lighting was made. For comparison with the earlier model the seasonal variation of the interval 40 demand about the annual mean is redrawn as curve (i) in figure 6.17(c). Curve (ii) is the same curve corrected for temperature and daylight effects. This curve shows that the model of equation 6.27 explains slightly more of the seasonal variation than the previous model (of equation 6.25).

A set of daily load curves corresponding to the actual internal temperatures used in the analysis above was generated from the model of equation 6.27. The result is shown in figure 6.19 which may be compared with the actual load curves of figure 6.4. There is considerable agreement between the two figures; however the model tends to over emphasize the temperature sensitivity, particularly in the morning.

While the demand-climate model described goes a considerable way toward explaining the behaviour of the demand from day to day it cannot be considered satisfactory. It has not been established behond doubt that it is a correct model of load behaviour. In particular the influence of other climatic variables, e.g. wind, cloudcover, found to be significant by other workers (14, 24, 30, 87) has not been investigated in the New Zealand situation. Such an investigation would be the basis for further research. Ripplay control of the water heating demand was employed in the particular load studied; the effect of this has not been determined.

It should be pointed out that in figure 6.17 (for daily interval 40) the sun was set on days 1-36 and 48-53. Consequently it is unlikely that cloud cover variations are a cause of all the residual variation observed in figure 6.17. Furthermore there was little observable correlation between this residual and actual wind velocity at 2000 hours. It is possible that the cooling effect of the wind has a lagged effect similar to actual temperature, although this has not been investigated. Thus the cooling effect of the wind on a body (building) could be calculated, using for example the relationship used by Davies (14); i.e.

$$\text{cooling value} = (\text{wind velocity})^m \; (T_b - T)$$

Where T_b = temperature of the body, i.e. the "comfort threshold" temperature,

T = ambient temperature, and

m = a constant; Davies suggests 0.5.

An "effective temperature" could then be calculated by subtracting the cooling value from the ambient air temperature. This "effective temperature" would then be used in the calculation of the internal temperature θ.

The utility of incorporating additional variables into the model should be considered. Though historical variations may be explained better it is difficult to forecast variables such as wind velocity with useful accuracy. Thus the uncertainty in the demand forecasts is unlikely to be reduced. A reduction in uncertainty is the object of incorporating additional variables.

All the models discussed so far have been developed for the purpose of forecasting short term load behaviour. Their application to forecasting is discussed in the next section.

6.5 Forecasting applications

6.5.1 Applications for models based on time series.

The three component and Fourier series models are based on time series approximations to the seasonal behaviour of the daily load curve. They do not explain day to day behaviour. Consequently neither can be used directly for forecasting demand over lead times of a few hours to a few days. The daily load curves generated by either model are those which would occur given average conditions. These models can, therefore, be used to generate "standard" load curves for any specified time of the seasonal cycle (i.e. of the year). "Standard" load curves form the basis for several short term forecasting methods as described in Chapter 4 and also in (37,73).

It is also possible to generate sets of load curves for entire seasonal cycles using the three-component-model. The normalized load curve is assumed to maintain the same shape on corresponding days in consecutive seasonal cycles. This means that appliance usage patterns and relative load composition are assumed to remain constant over the desired number of seasonal cycles. A set of normalized daily load curves for each cycle is generated from equation 6.2. Each curve is scaled by the appropriate daily energy usage obtained from equation 6.11. It is only necessary to determine the mean and maximum values of daily energy usage in the seasonal cycles for which sequences of load curves are desired. These values may be estimated using standard long term energy usage forecasting techniques, e.g. simple extrapolation. A sequence of load curves similar to that of figure 6.8 is generated.

The Fourier series model may also be used to generate load curves for entire seasonal cycles. However it is necessary to estimate the Fourier coefficients in each interval of the day for each seasonal cycle for which the load curves are to be generated. This involves considerably more labour than is the case for the three component model.

A further, and important, application for these models is the generation of data for system planning studies. These include reliability investigations, the optimum use of available energy resources, etc.

6.5.2 Applications for demand-climate models

A demand-climate model contains sufficient information to reproduce the behaviour of the load on any particular day, given a knowledge of the weather etc for that day. The quality of the reproduction is determined by the quality of the model. Average, or smoothed, behaviour can be reproduced by using average weather statistics in the model. Short term demand forecasts are obtained from these models by using forecasts of the weather for the desired lead times.

The major sources of uncertainty in demand forecasts obtained from demand-climate models are;

(a) errors in estimating the model parameters, i.e. a_{Bt}, $a_{\Theta t}$, and

(b) errors in forecasting the relevant weather variables, e.g. internal temperature of buildings.

The first source is the result of vagaries of human behaviour and of modelling inaccuracies, e.g. not all the relevant variables are included in the model. The amount of uncertainty contributed by this source can be estimated during the regression analysis but it is difficult to reduce. A more comprehensive model may reduce the uncertainty but total elimination requires extensive centralized control over all aspects of energy usage, i.e. ripplay control applied to all appliances. In general the standard errors of the estimates of $a_{\Theta t}$, $t = 1, \ldots, M$ obtained from the regression analysis of section 6.4.3 were about ten percent of the estimate, i.e. $\mathrm{var}(a_{\Theta t}) \approx 0.1\ \hat{a}_{\Theta t}$.

In the course of an investigation into the usefulness of ambient air temperature forecasts in New Zealand Maunder (76) obtained standard deviations of about 2°F for the distribution of forecasting errors. These experimental forecasts were made at about 1600 hours daily for lead times up to 24 hours over a period of three weeks in July 1970. A standard error of this magnitude corresponds to a standard error in the demand forecast of about 8 MW at the time of the daily peak using the results of, for example, figure 6.18; this is about 4% of the annual maximum demand for this particular supply authority. In the temperature forecasting study quoted it was found that extremes of temperature were the most difficult to predict. However, because the internal temperature of buildings is effectively smoothed, such extremes will only affect the demand significantly if they are prolonged.

It is difficult to forecast the weather over lead times greater than one or two days. While it is possible to predict the type of event, e.g. rain or snow, over longer lead times, the timing and magnitude are more difficult propositions (67, 70). Consequently mean weather conditions for the particular time of the year must be substituted for weather forecasts over lead times of three or more days. The uncertainty in the resulting forecasts may be determined from the historical distribution of temperature about the long run seasonal means; e.g. figure 6.20 which shows the weighted (according to regional energy usage) temperature for the North Island of New Zealand during winter. This distribution has a standard error of about 2.7°F; this corresponds to a national maximum demand forecast standard error of about 60 MW, using the one percent load change per degree F. factor quoted by Mackenzie (9). To be 95% confident of being able to meet the demand in such circumstances would require spare capacity of at least 120 MW, assuming the demand-climate model contains no errors. This exceeds, by about 20%, the desirable amount of spare capacity determined in Chapter 3; i.e. spare capacity equivalent to the capacity of the largest generating unit. Hence, unless temperature forecasting uncertainty can be reduced by about half it will be necessary to strive for the greatest possible accuracy in demand models.

All the models described have been based on the daily load curve for a particular day of the week. They may be extended to all days of the week by introducing day-of-week factors, e.g. Davies (14). Alternatively the models may be applied to the weekly load curve which is simply the load curves for each day of the week in sequence. This approach has been used with some success by Christiaanse (10) with a forecasting scheme based on general exponential smoothing.

6.6 Conclusions

Three models of the short term behaviour of the load have been developed. Two of these model the seasonal changes in the daily load curve. Using these models it is possible to generate "standard" load curves for any particular time of the seasonal cycle, or to generate a set of load curves for a complete cycle for use in system planning studies. Both of these models represent load behaviour in average environmental conditions.

The third model is of the demand, temperature and daylight illumination relationship. It has been shown that the demand is more closely related to the internal temperature of buildings than to the ambient air temperature. The available evidence suggests that the thermal time constant of dwellings in New Zealand is about 8 hours. The demand-internal temperature relationship is shown to be approximately linear with temperature above and below a "comfort threshold" of about $60^{\circ}F$. Daylight illumination affects the seasonal behaviour of the demand only at those times of the day between the earliest and latest times of sunrise and sunset. Lighting appears to contribute about 100 W per domestic consumer at the times of the day studied. The utility of incorporating explanatory variables which are difficult to forecast is questioned.

It has been shown that if demand forecasts over lead times of about 24 hours are to even approach the desirable accuracy determined in Chapter 3 then the model uncertainties must be reduced to a minimum as must temperature forecast uncertainties. The vagaries of human behaviour impose a lower limit on the accuracy. This limit cannot be lowered unless extensive, and expensive, load control measures are introduced.

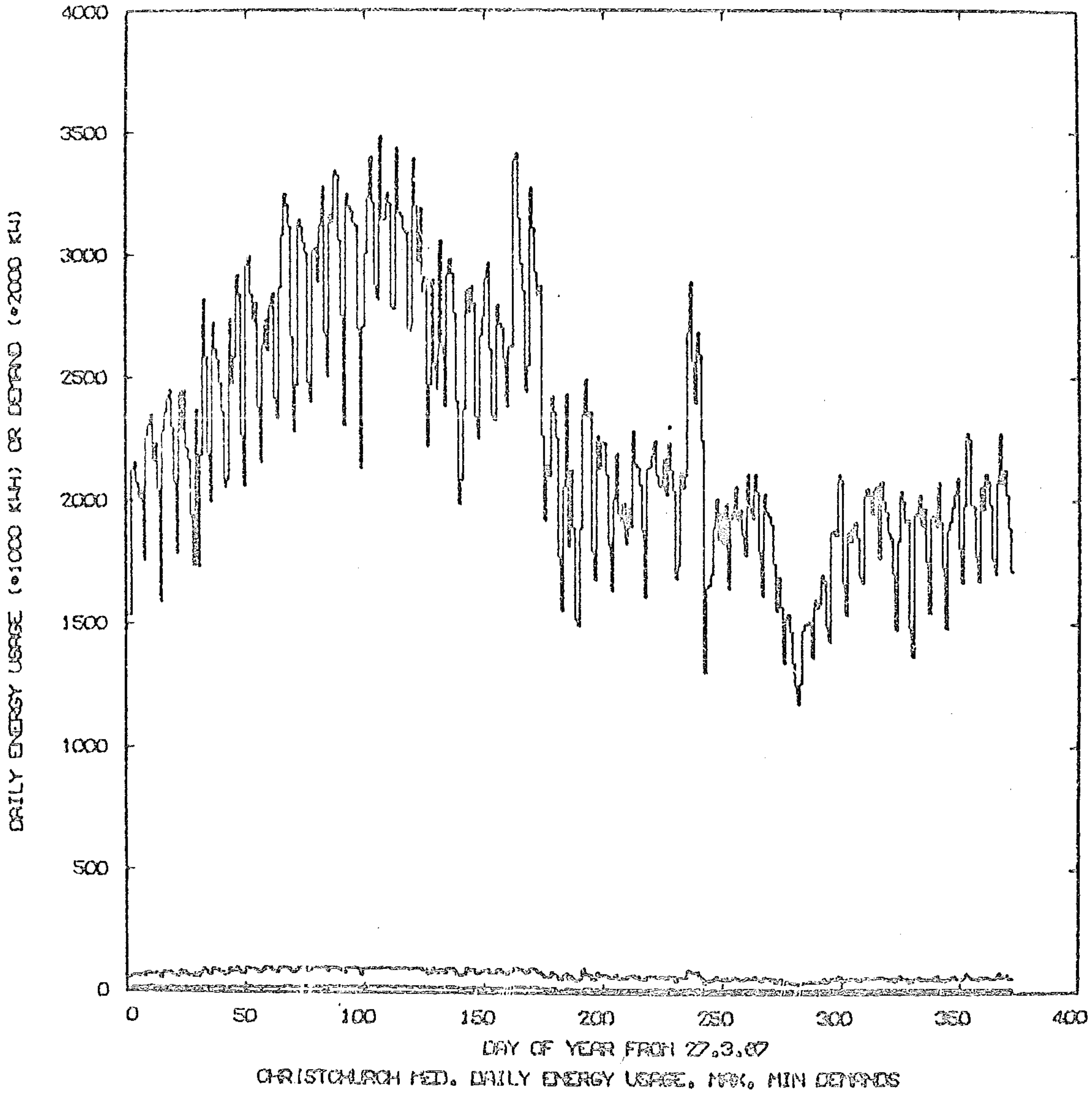

4000
3500
3000
2500
2000
1500
1000
500
0
0
50
100
150
200
250
300
350
400
DAILY ENERGY USAGE (*1000 KWH) OR DEMAND (*2000 KW)
DAY OF YEAR FROM 27.3.67
CHRISTCHURCH MED. DAILY ENERGY USAGE, MAX, MIN DEMANDS

Figure 6.1

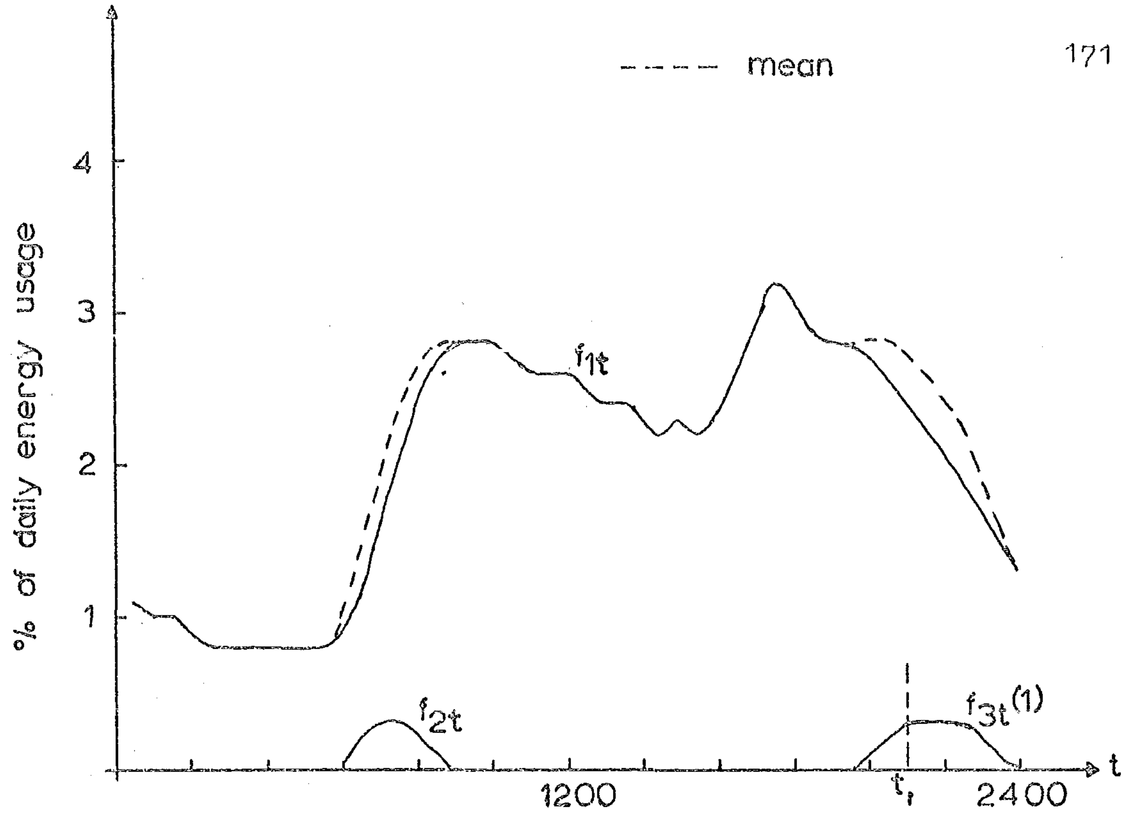


Figure 6.2 Initial component load curves.

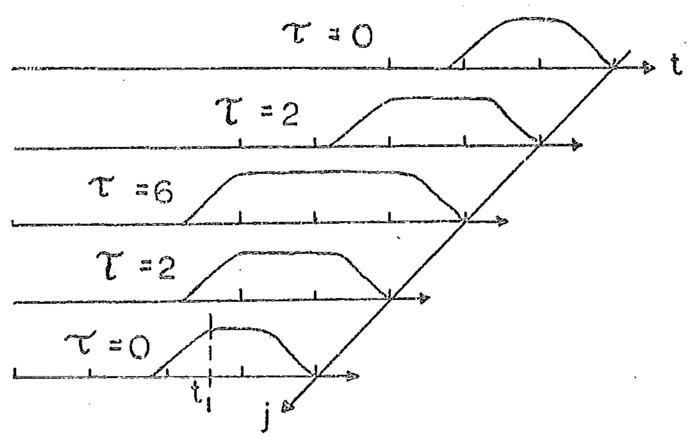


Figure 6.3 Shape of evening component with j

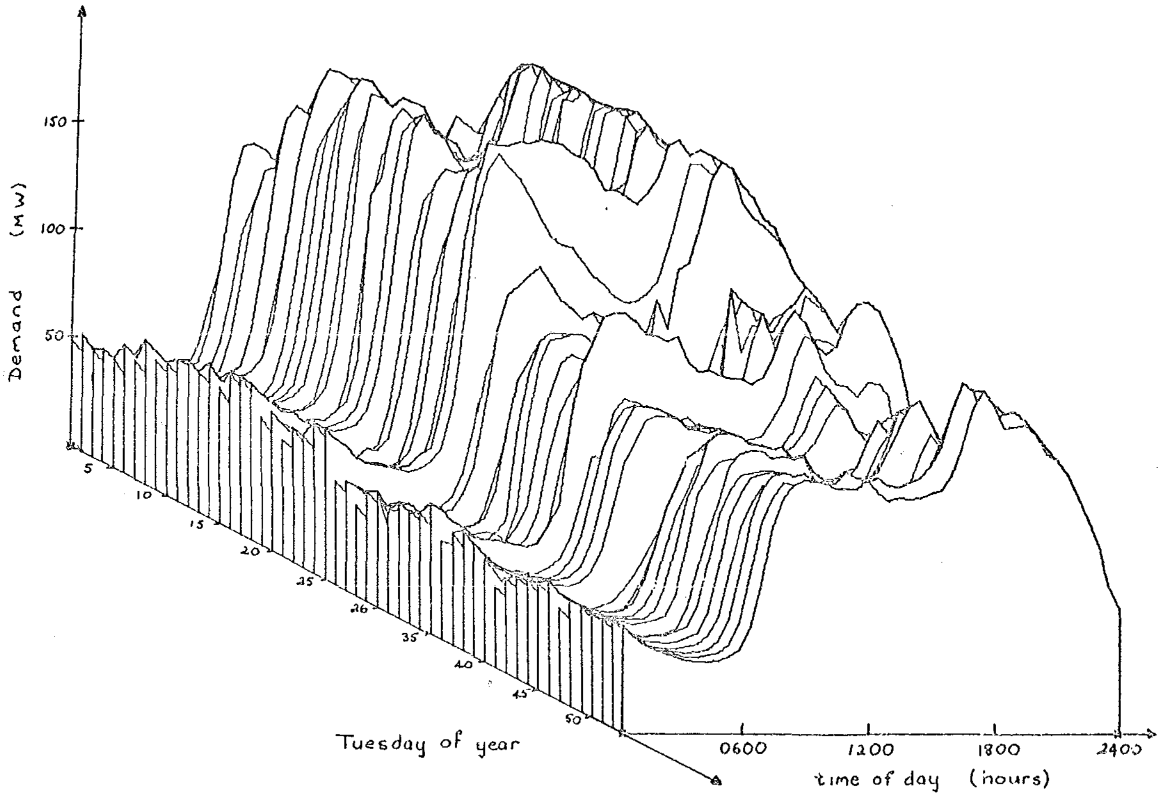


Figure 6.4 Load curves for Tuesdays from 28.3.67: Christchurch MED.

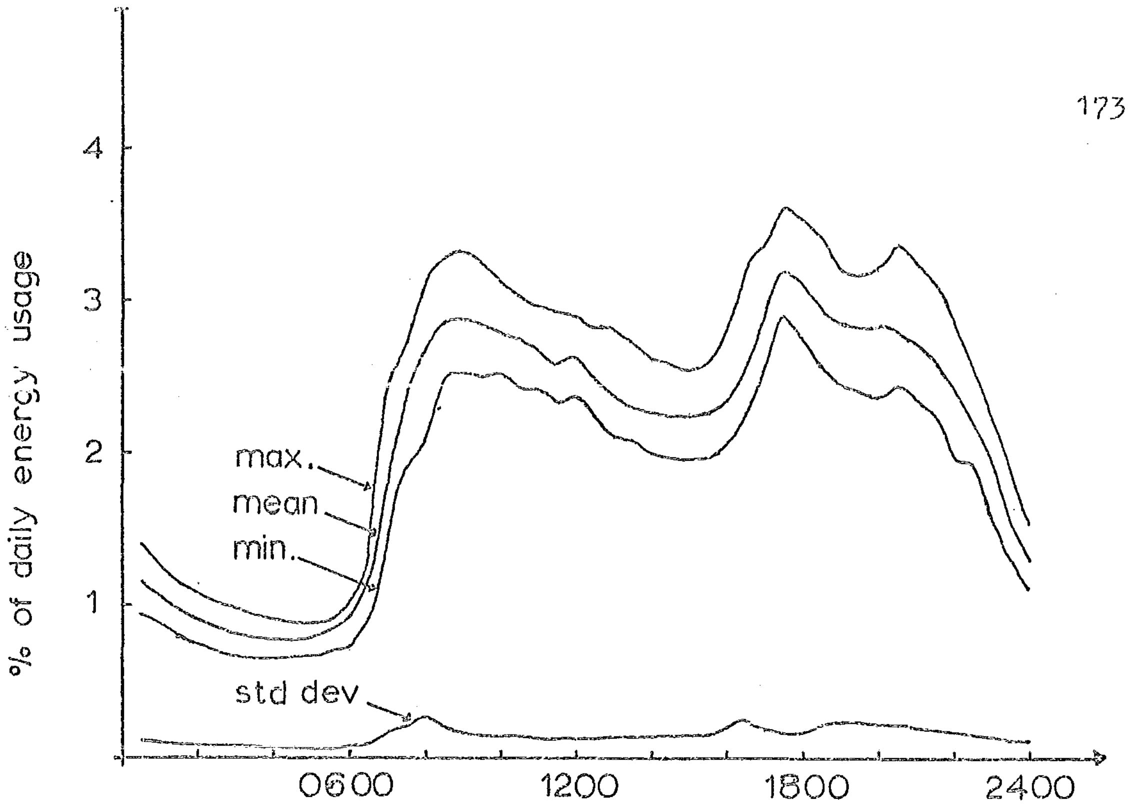


Figure 6.5 Statistical summary of normalized daily load curves

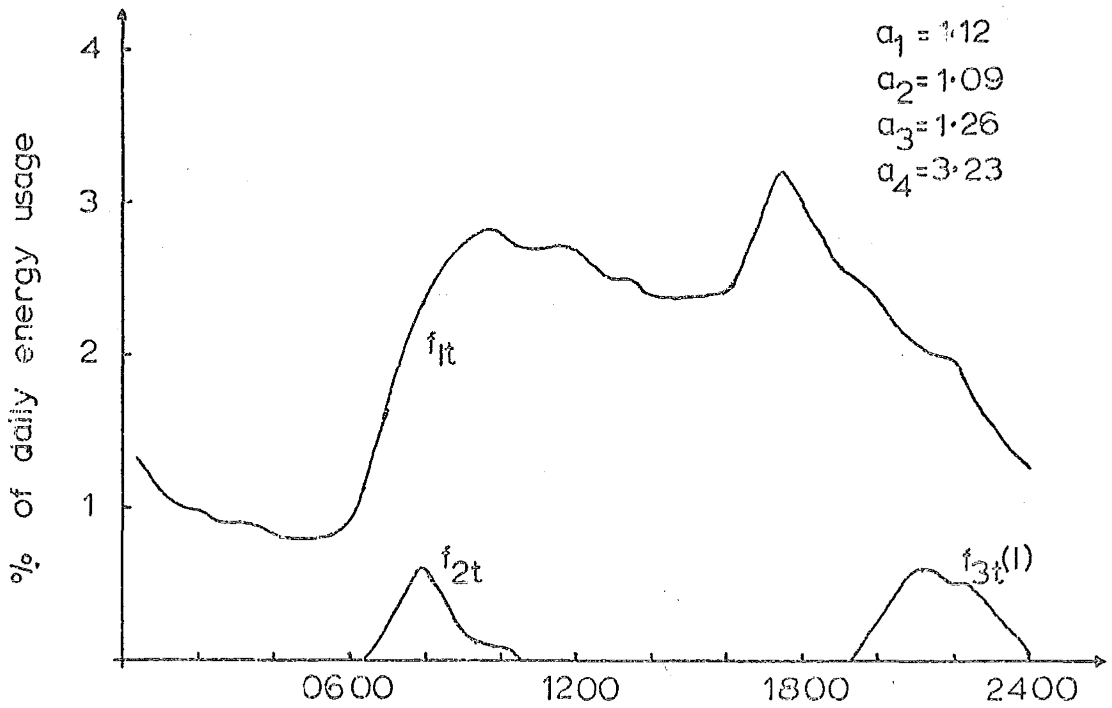


Figure 6.6 Final component load curves and scale factors

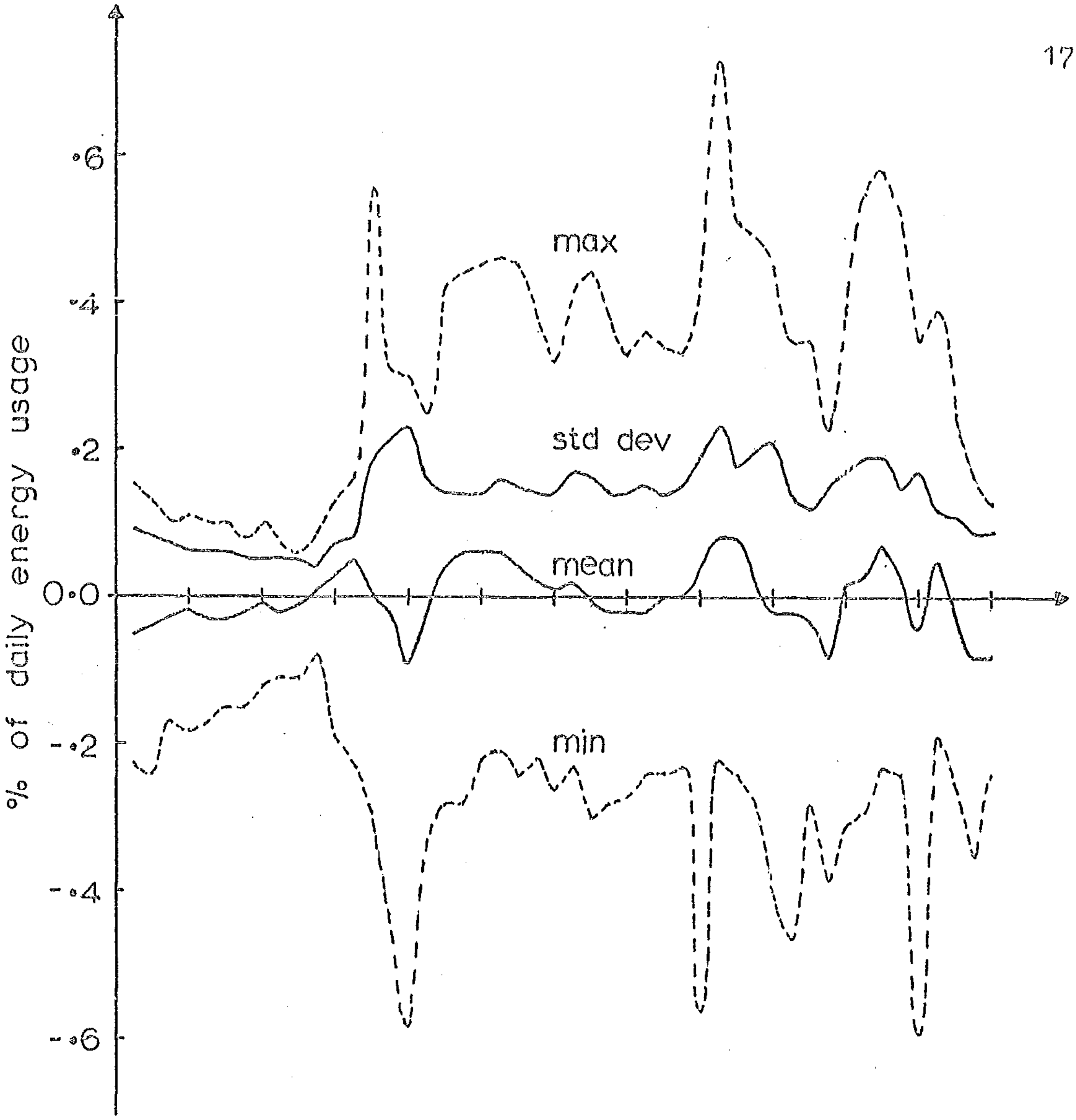


Figure 6.7 Statistical summary of mismatch between three component model and raw data.

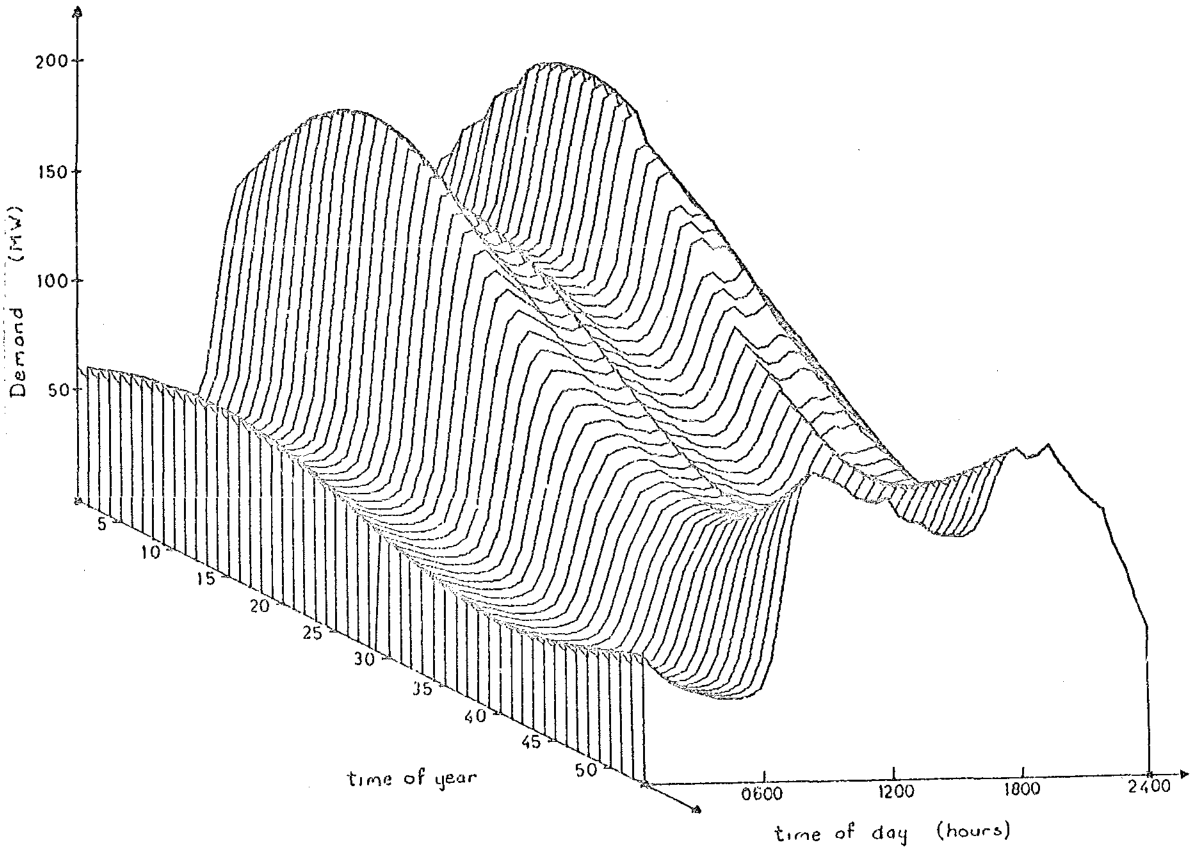


Figure 6.8 Load curves generated by three component model.

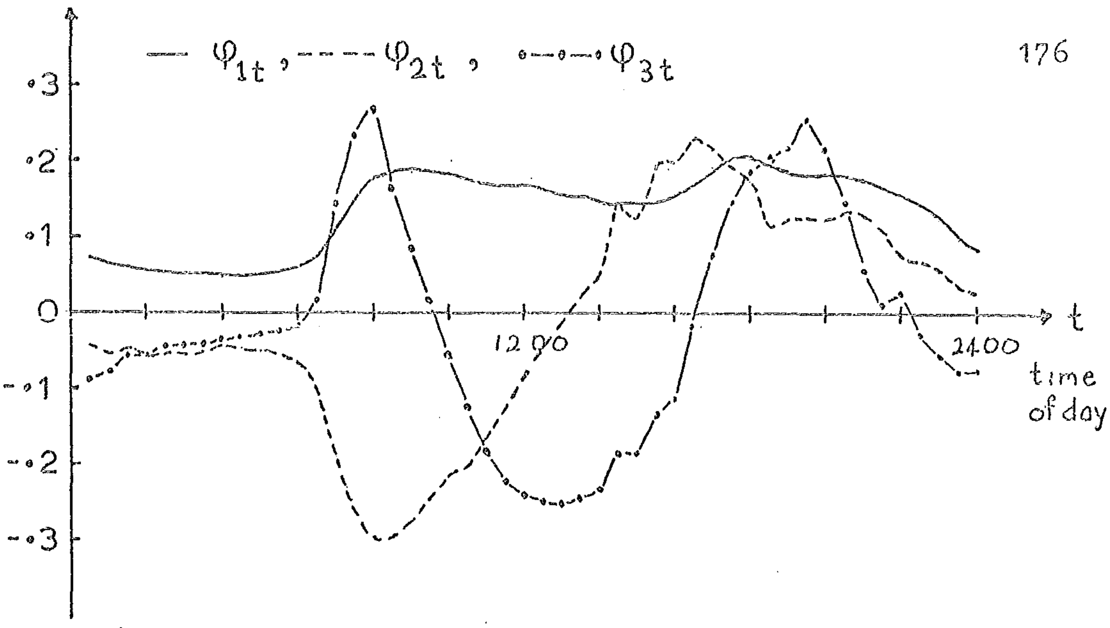


Figure 6.9(a) Eigenvectors of covariance matrix

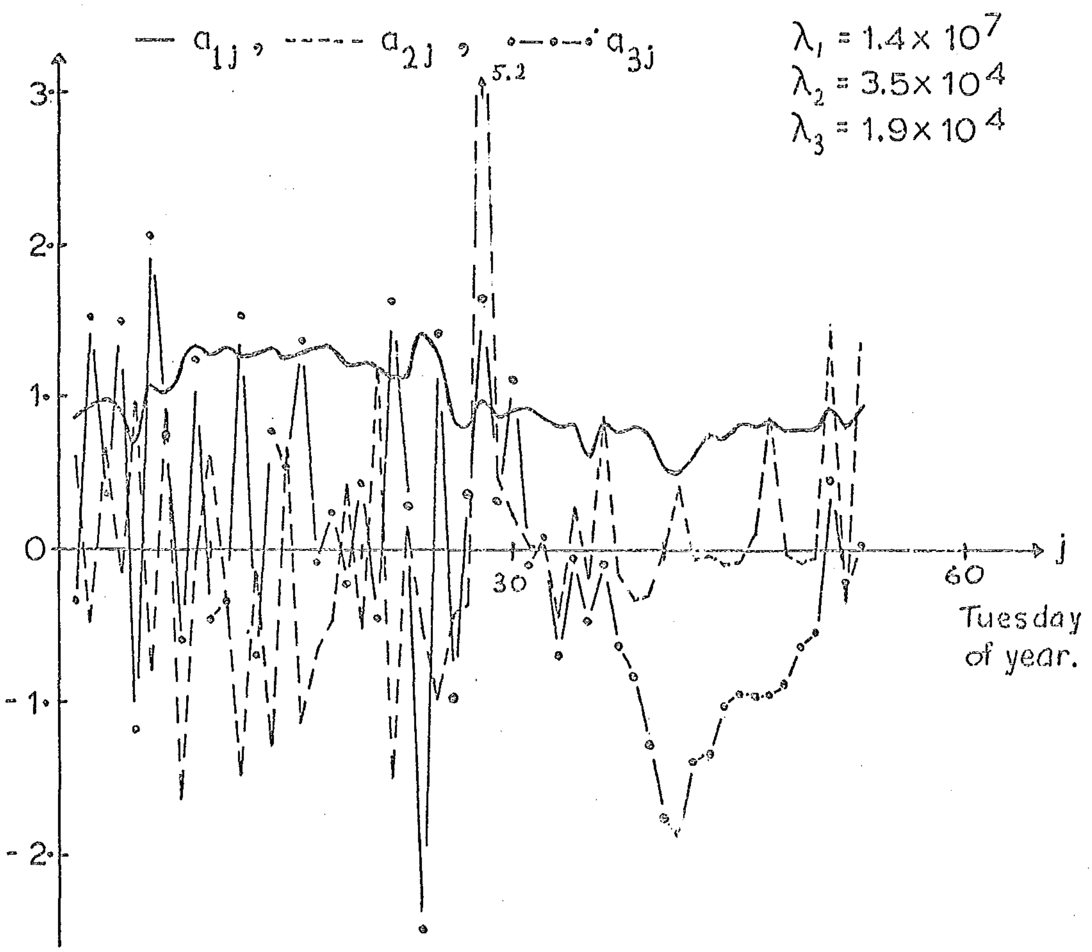


Figure 6.9(b) Spectral coefficients for Tuesday load curves

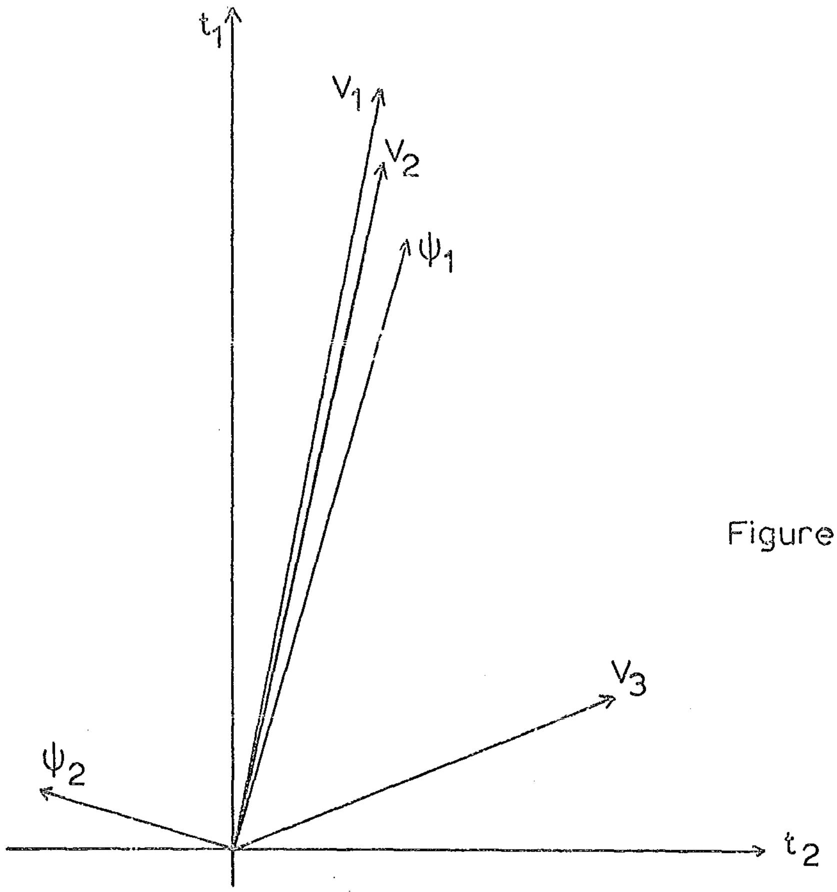


Figure 6.10(a)

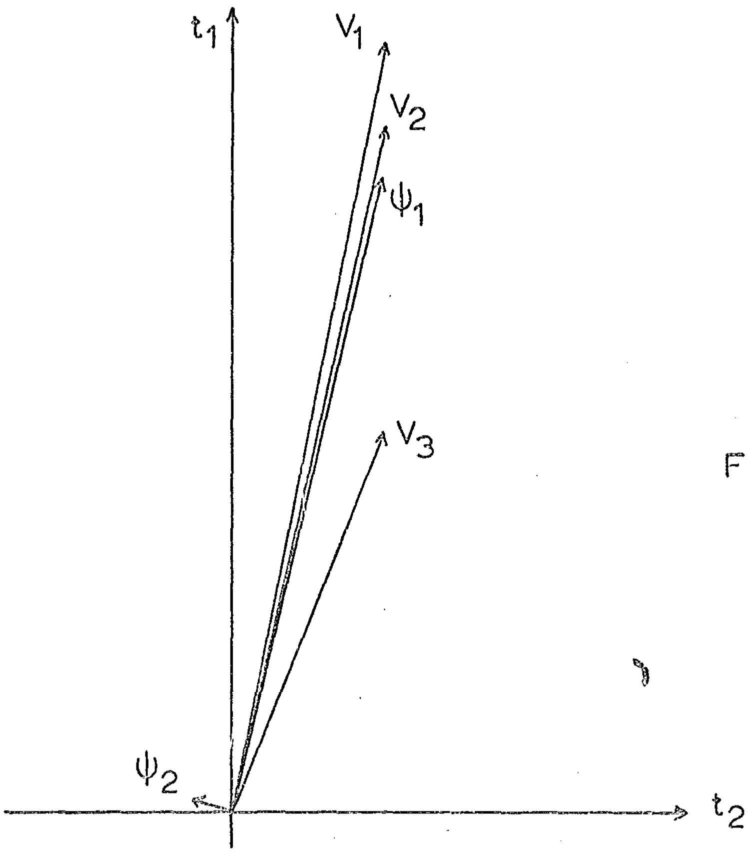


Figure 6.10(b)

Operation of the Karhunen Loeve Expansion

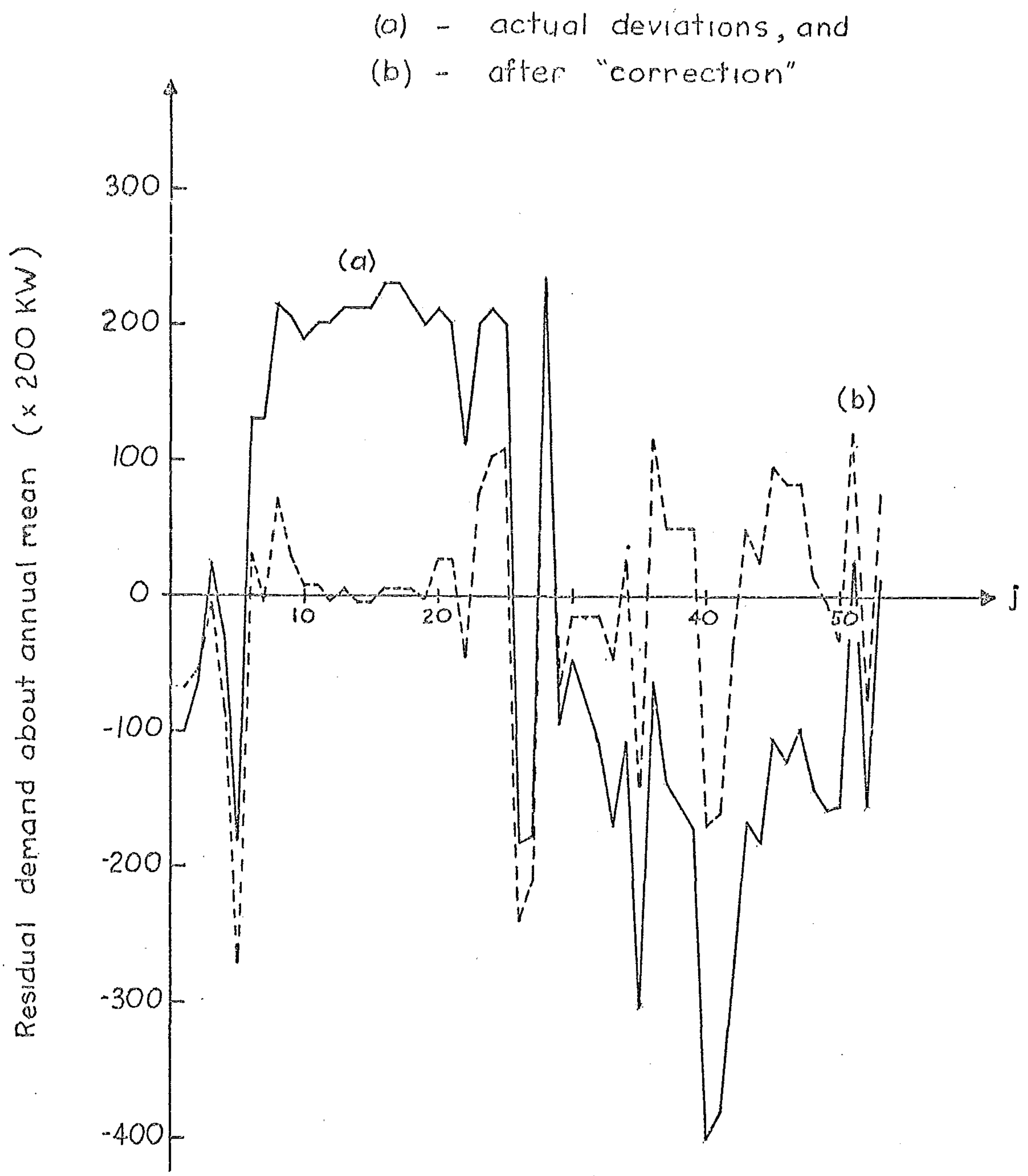


Figure 6.11 Fourier series approximation to the integrated demand in interval 36 on Tuesdays from 27.3.67

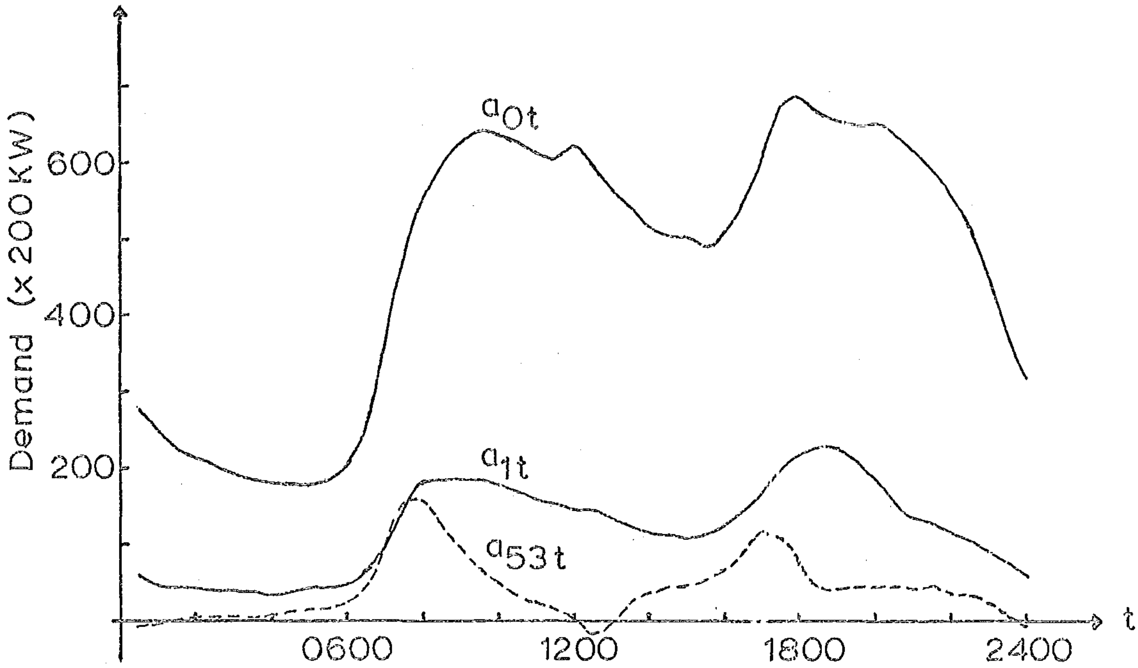


Figure 6.12 Fourier components from one year's demands.

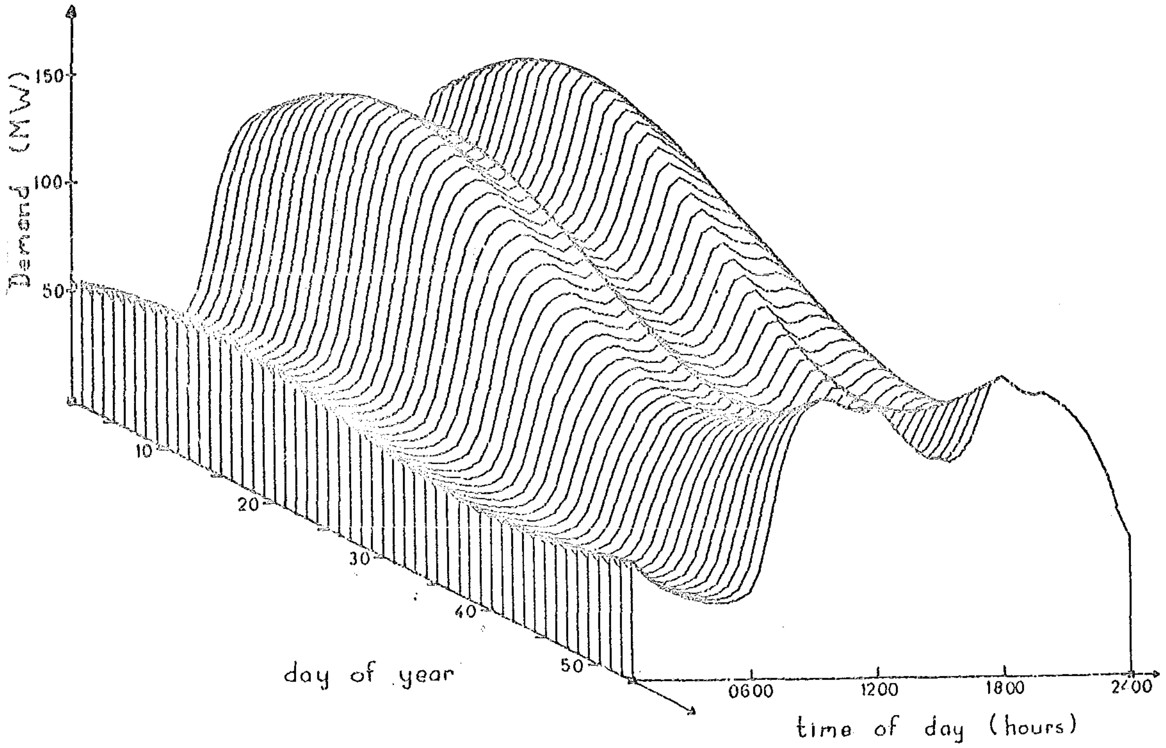


Figure 6.13 Load curves generated by Fourier series model

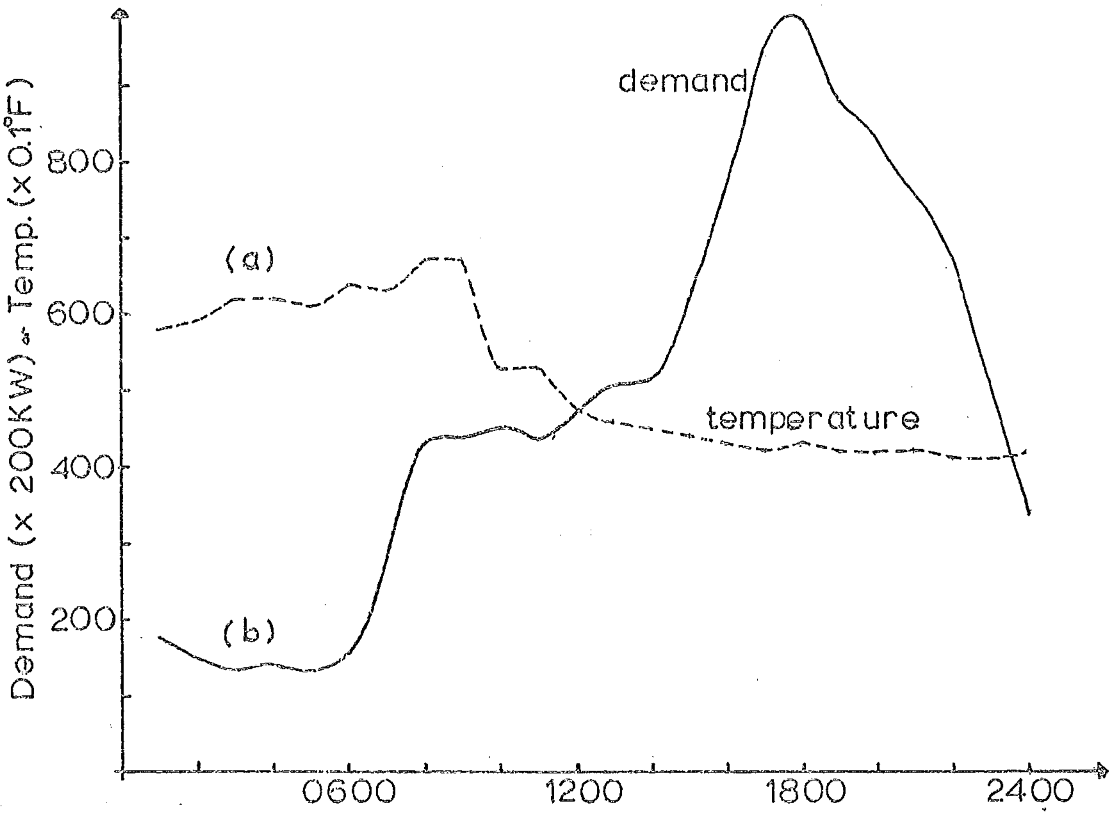


Figure 6.14 Demand and temperature on Tuesday 3.10.67

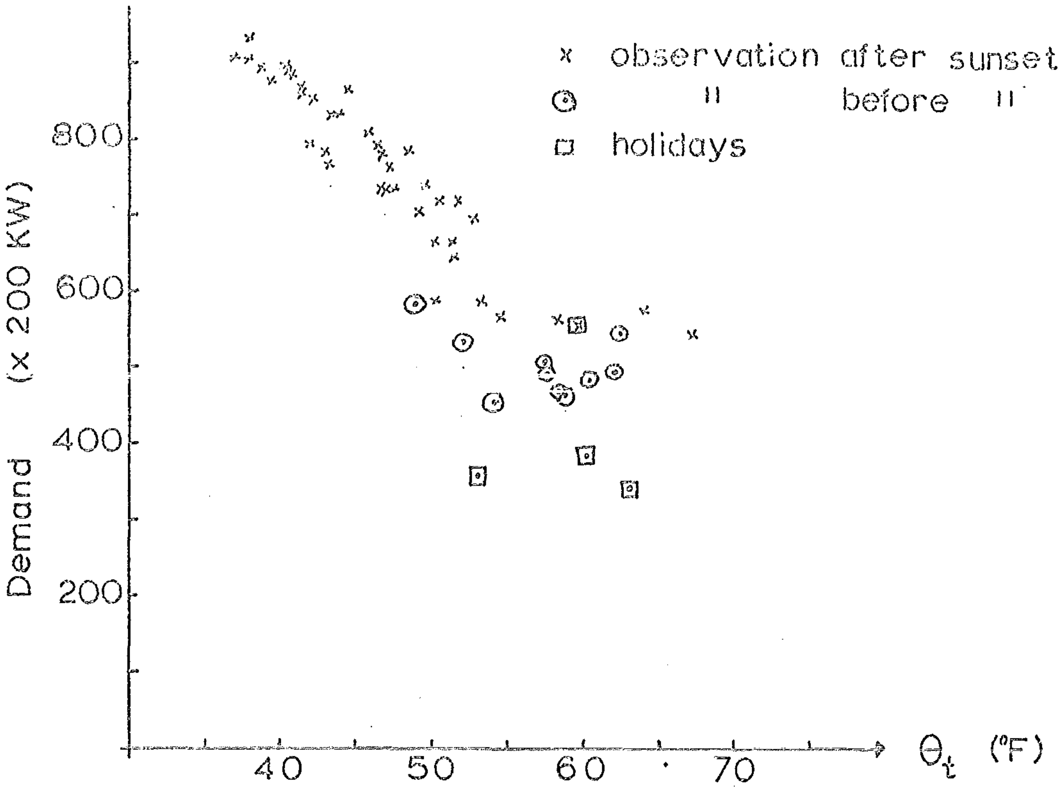


Figure 6.15 Demand and lagged temperature for interval 40

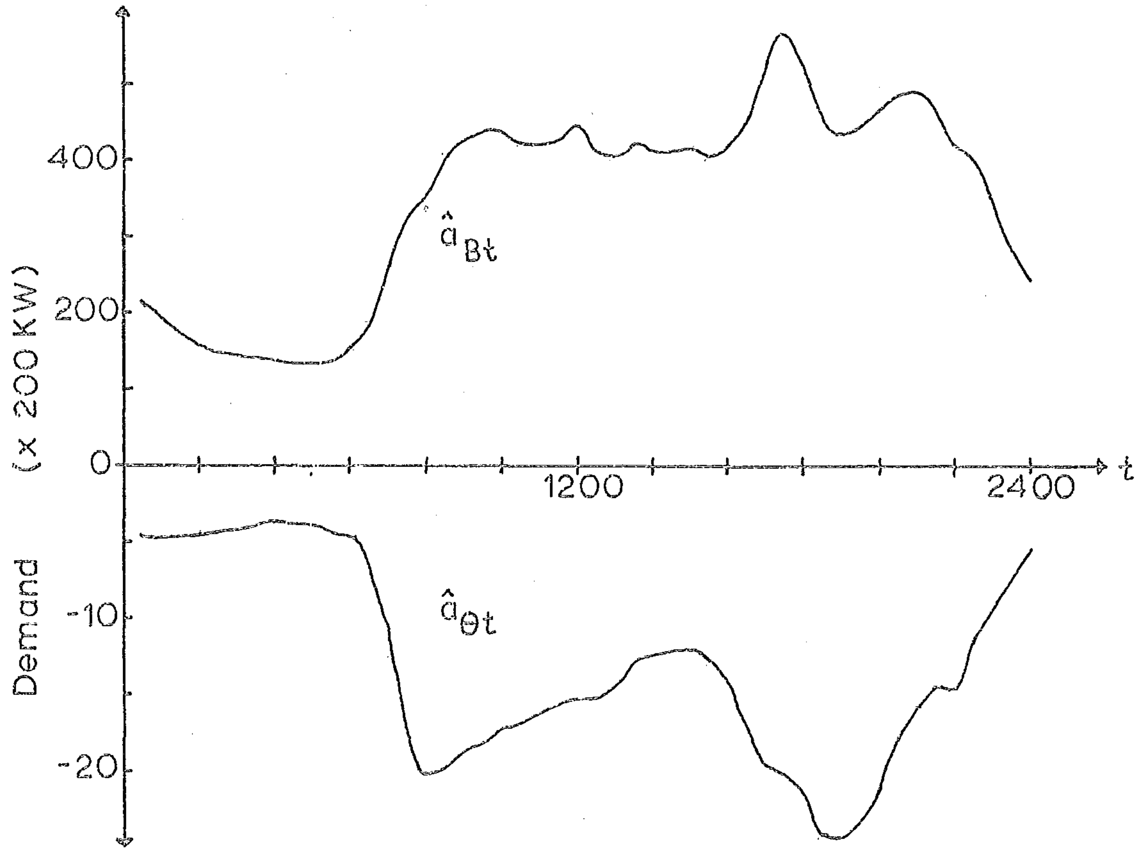


Figure 6.16 Estimated coefficients for demand-lagged temperature model

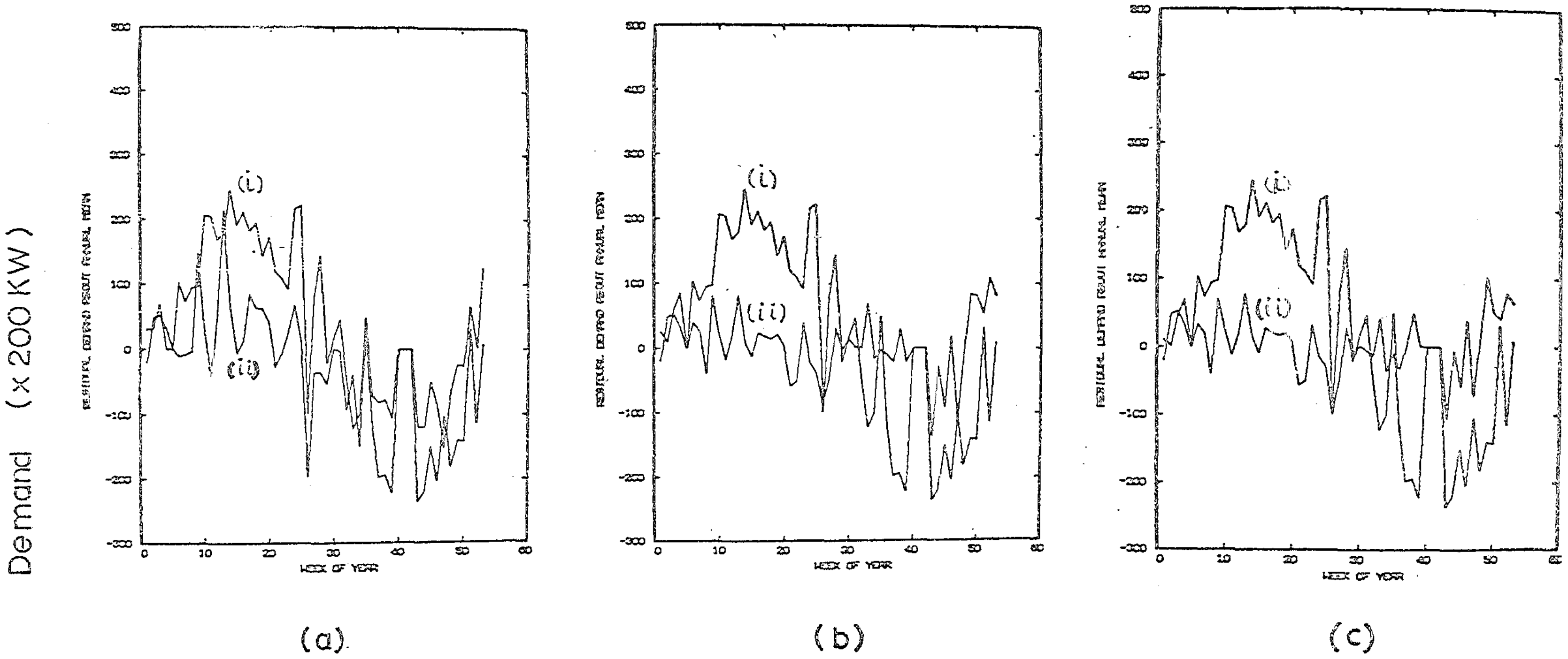


Figure 6.17 Residual demand about the annual mean in daily interval 40 approximated by three demand-climate models

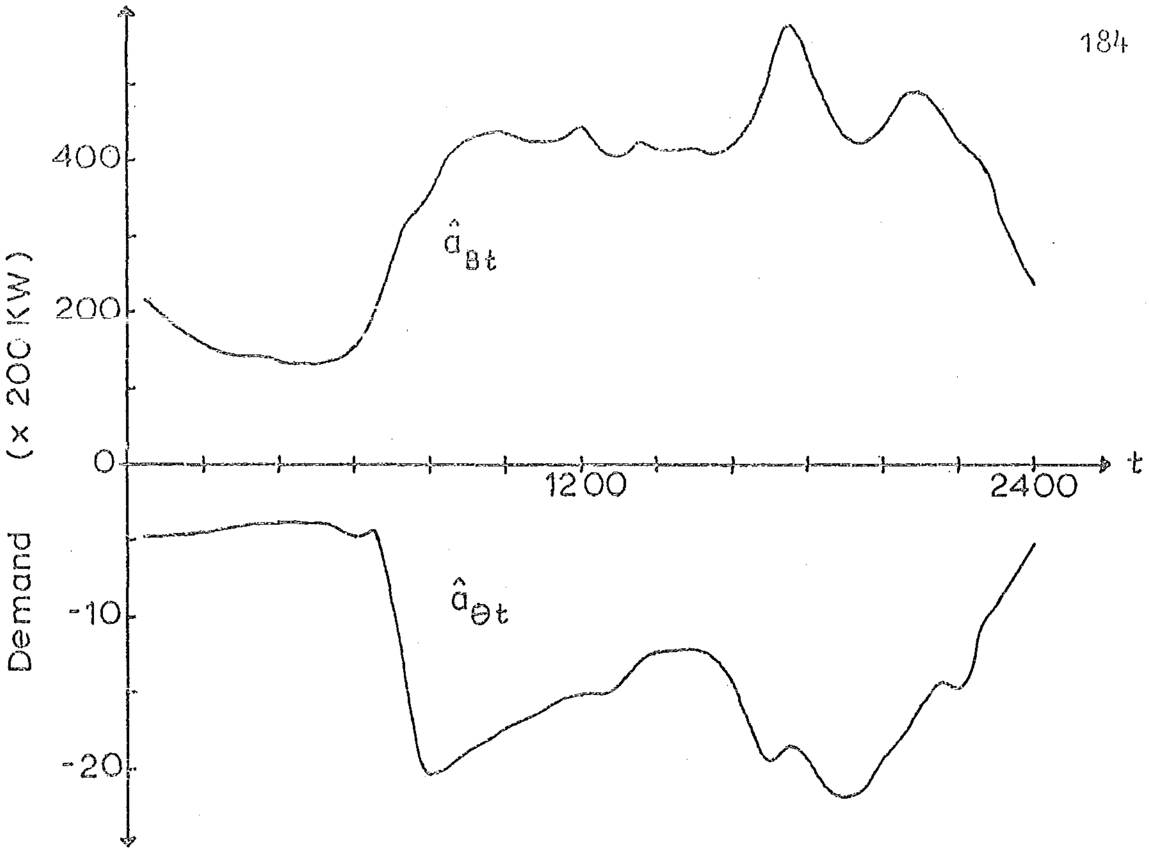


Figure 6.18 Estimated coefficients for demand lagged temperature daylight model

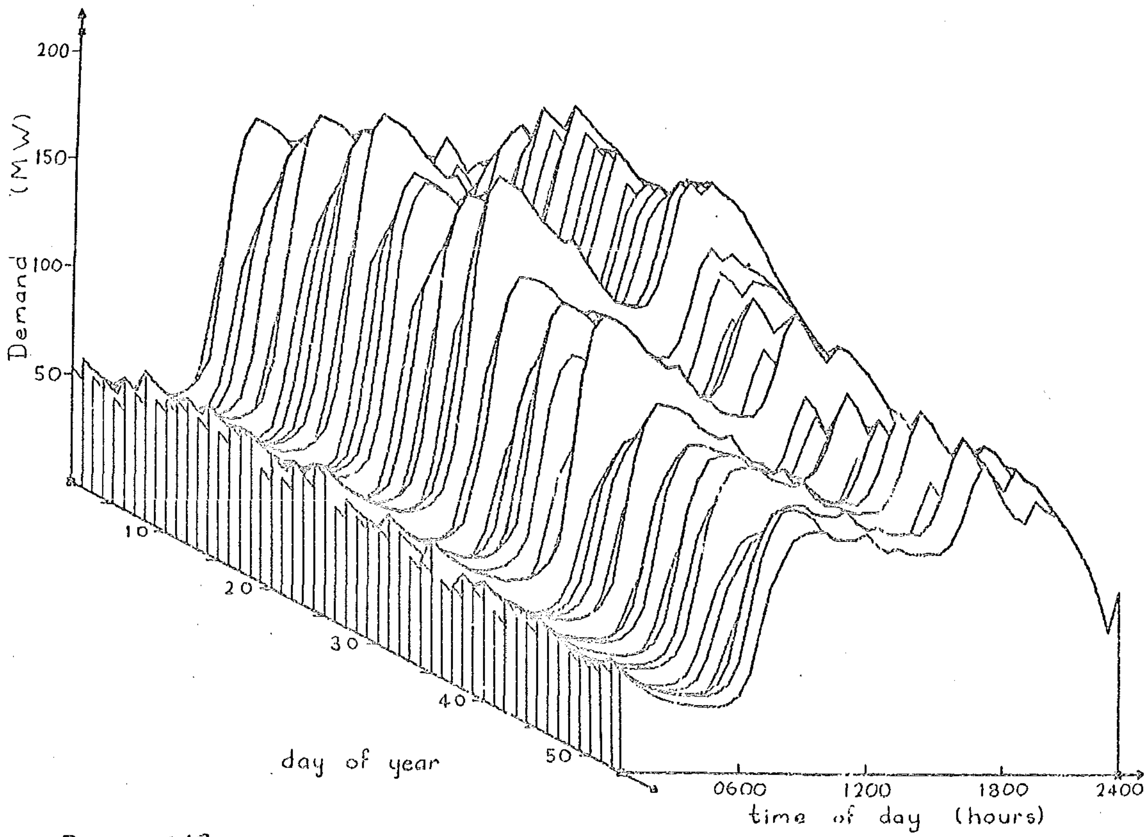


Figure 6.19 Load curves generated by demand-weather model.

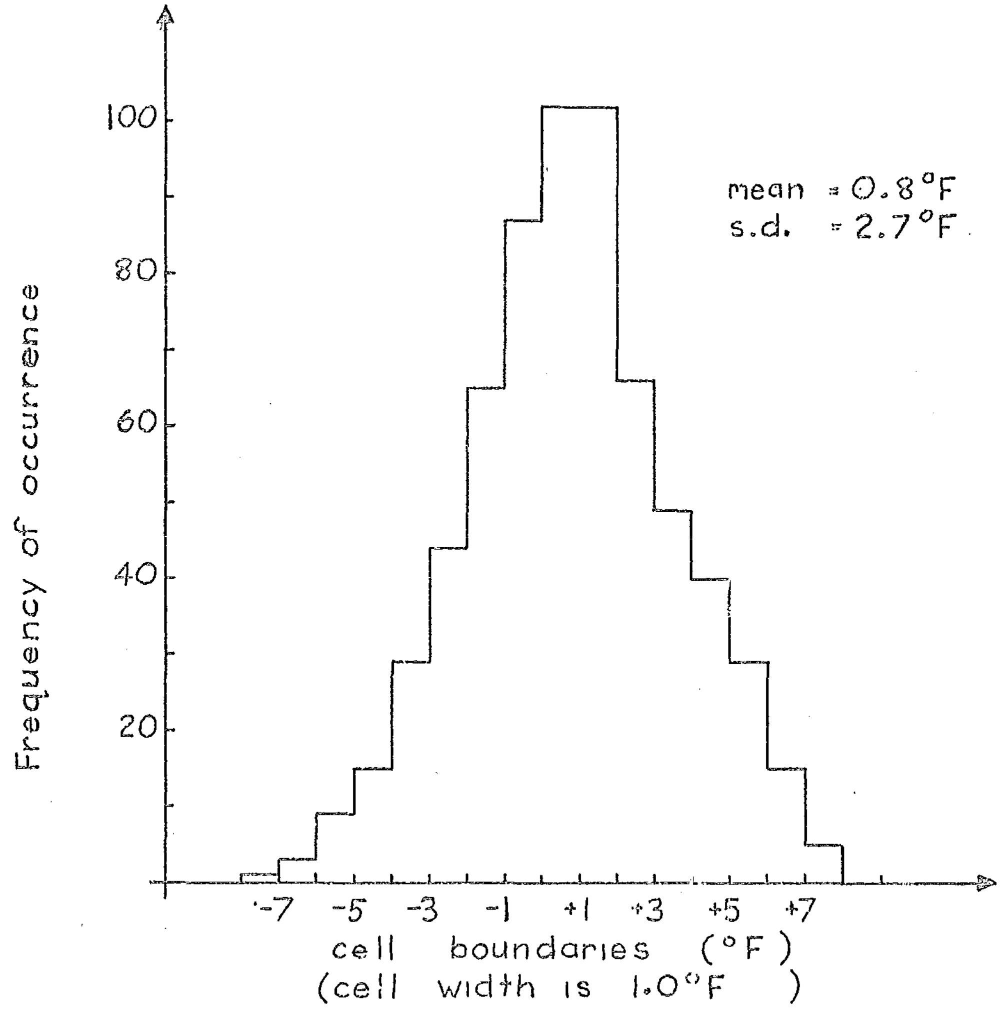


Figure 6.20 Deviation of North Island weighted temperatures from mean ; June and July 1960 - 1970

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS

7.1 Conclusions

The principal achievements of this thesis, and their significance, are summarized in this section.

In chapter 2 the load modelling problem is formulated in two parts;

(a) modelling of the process governing the growth of the load's potential for energy usage (increases in the number of appliances owned by consumers) and,

(b) modelling the way this potential is used (the way consumers use their appliances).

This formulation emphasizes the distinction between the long term and short term forecasting problems. The long term forecast is mainly concerned with the way the number of appliances which consumers own increases. Short term forecasting is concerned with how these appliances are used.

A criterion for determining whether a forecasting method is sufficiently accurate for its proposed application is derived in chapter 3. Sufficient accuracy is defined in terms of the cost of providing spare capacity to ensure the risk of not meeting the demand, due to uncertainty in the forecast, is less than some value specified by management. It is argued that competing methods for an application should be compared on the basis of the amount of uncertainty in the forecasts (at some given confidence limit) and not on the basis of achieved forecasting error (i.e. forecast value minus actual value). Thus improvements to forecasting accuracy can be costed and compared with the cost of implementing the method. This criterion has been used to show that there is no advantage in reducing the uncertainty of forecasts below limits determined by the unreliability of plant.

A model of the domestic load growth process, based on its structure, has been postulated in chapter 5. The model has two parts;

(a) an explanation of growth in domestic consumer numbers based on the population birth and death process and,

(b) an explanation of growth of an individual consumer's potential for energy usage based on the way consumers accumulate appliances.

This model enables the simulation of the domestic load growth process, starting from known information, e.g. population statistics. Forecasts over lead times of ten to fifteen years can be made without needing to forecast the independent variable , i.e. population. The effect on the load growth of controls on, for example, household expenditure can also be studied with this model; such studies are difficult to perform with models based on fitting a mathematical function to historical data.

An elementary model of non-domestic load growth is developed in chapter 5. This relates growth of energy usage to growth in the demand for the goods and services provided by non-domestic consumers. The model recognizes that non-domestic electric energy usage is derived from the demand for goods and services, and that other sources of energy may be substituted for it in production processes. At present a lack of data prevents the proper evaluation of both this model and the one explaining domestic load growth.

It has been demonstrated, in chapter 6, that the seasonal changes in daily load curve shape can be modelled satisfactorily by three daily load curve components. The shape of these components may be obtained by artificial means or by professional judgement. The contribution of each component to the overall shape is assumed to vary sinusoidally with the time of year. This model is particularly suitable for generating sequences of daily load curves for an entire seasonal cycle, knowing only the average and expected maximum (or minimum) daily energy usage for that cycle.

Also in chapter 6 a comprehensive demand-temperature model is applied in the New Zealand situation (which differs from most overseas situations because of the large, 55%, domestic portion of the total demand). This model is based on the one developed by Davies (14) which incorporated a complex non-linear relationship between demand and the internal temperature of consumers premises. In this thesis a considerably simpler piece-wise linear relationship has been used. Also the model has been applied to all daily intervals rather than only to those in which maximum demands occur, as done, for example, by Davies (14) and Hinemann (30). Experiments with this model show, conclusively, that consumer demands respond to a weighted sum of past temperatures and not simply to the present ambient temperature. The presence, or absence, of daylight accounts for a small amount of the variation in demand not explained by the sum of the weighted temperatures.

The review of the state of the art has shown that extrapolation of mathematical models fitted to historical data forms the basis of existing long term forecasting methods. Though simple in concept, and to implement, these methods have two major disadvantages;

(a) the assumption that the growth process will continue to behave over the lead time of the forecast as it has during the relevant period in the past, and

(b) the large amount of uncertainty in the forecasts due to a paucity of data.

It is also difficult to assess the value of proposed load control measures or to assess the implications of changes in load composition using these models. The forecasts obtained are, however, useful approximations to the future provided the implications of these inherent disadvantages are understood.

Short term (daily based) demand forecasting is at present based on either demand-weather models or methods involving the scaling of standard load curves. Demand-weather models offer the possibility of anticipating demand changes from changes in the weather. However accurate forecasts of all relevant weather variables are required and these are difficult to obtain, except in some special cases (69). The scaling methods in effect extrapolate historical demand behaviour; there is no inherent ability to anticipate changes although a recent model based on state estimation (37) considers rate of change information. Scaling methods are simpler to use than demand-weather models as they use information which is readily available to the system operator. Provided the weather is changing only slowly from day to day and within days these methods are capable of giving good results.

7.2 Recommendations

7.2.1 Forecasting procedures.

The methods based on extrapolation of historical data are recommended for a first approximation to long term maximum demands or energy requirements. Where little quantitative information on the structure of the load growth process is available extrapolation methods must be used. To obtain more data for estimation of the uncertainty in forecasts it is suggested that monthly records of, say, energy usage be corrected for seasonal variation; e.g. (11, 26, 90, 91) and models fitted to the resultant "smoothed" monthly series. If the available evidence indicates that the assumptions inherent in these methods are valid there is little need for complex models unless control actions are contemplated.

In situations where the load composition is changing significantly (as in New Zealand with increasing industrialization), or an examination of load control measures is desired structural models of the load growth process are recommended. It is essential in all cases that the uncertainty in the forecasts is estimated. This determines how much additional capacity (and capital investment) is required to obtain reliable supplies of electric energy.

Scaling methods, based on the weekly load curve, are recommended for short term demand forecasting applications. The standard load curve may be scaled on the basis of deviations of historical demands from the standard, e.g. (73), or on the basis of weather variations about the long run seasonal means (where satisfactory weather forecasts are available).

7.2.2 Further work

The acquisition of sufficient suitable data to enable the evaluation of the structural models proposed in chapter 5 is recommended as the next step in the modelling and simulation of the load growth process. It is anticipated that appliances will need to be metered, for a selected sample of consumers, to determine typical appliance usage factors. A survey of domestic consumers' preferences for appliances and of their accumulation strategy (on the lines discussed in chapter 5) will also be necessary to confirm that growth in a consumer's load occurs as postulated.

Further analysis of the contribution of electric energy to industrial and commercial output is recommended and in particular the relationships between energy and labour inputs. There is considerable scope for extension of the model of non-domestic energy usage by incorporating more comprehensive models of the population's demand for goods and services, such as those developed in economic studies (20, 92).

It is also recommended that further investigation of the role of the weather in determining the way consumers use appliances be carried out, particularly on the effects of cloud cover, rainfall and wind (both direction and velocity). These variables are believed to have second order effects, e.g. wind increases air flow through dwellings thus reducing the effective thermal time constant and making demand more sensitive to temperature change (the magnitude of this effect is also dependent on the orientation of buildings with respect to wind direction as well as velocity). Such a study, without the ability to accurately forecast the weather, will be of little direct benefit to short term forecasting but will enable

(a) improved estimation of the amount of uncertainty in short term forecasts, and

(b) "correction" of historical records to standard weather conditions to allow examination of the true nature of load growth.

Short term fluctuations in gas demand have been found to be more sensitive to deviations of weather conditions from the seasonal average than to absolute temperature (93). This implies that consumers' appliance usage habits change as people become acclimatized to different conditions. It is suggested that the demand for electric energy be examined along these lines also. This examination would have the aim of reducing the unexplained variation in demand.

APPENDIX A

Derivation of expected deficiency of demand

In Chapter 3 the expected deficiency of capacity is defined as

$$E\left\{u\right\} = \int_0^\infty u\ p(u)\ du \tag{A1}$$

where u = deficiency of capacity, $\geqslant 0$,

$p(u)$ = probability that a deficiency, u, will occur,

and

$$\int_0^\infty p(u) = 1$$

Using the notation of Chapter 3 and assuming normally distributed uncertainty in the demand forecast, equation 3.10 may be written as

$$\text{RISK}_2 = \sum_{j=1}^{n'} \left[\int_{G_j}^{\infty} \frac{1}{\sigma_v \sqrt{2\pi}} \exp\left\{-\frac{(L-D_f)^2}{2\sigma_v^2}\right\} dL \right] \sum_{q=1}^{Q_j} \prod_{i \notin s_{qj}} p_i \prod_{i \in s_{qj}} (1-p_i) \qquad \ldots\text{(A2)}$$

A capacity deficiency occurs when L, the load on the power system, exceeds the available generating capacity, G_j ; giving

$$\begin{aligned} u &= L - G_j && \text{for } L > G_j \\ &= 0 && \text{for } L \leqslant G_j \end{aligned} \tag{A3}$$

The probability of a deficiency occurring is then

$$p(u) = 0 \qquad \text{for } L \leqslant G_j$$

$$= \frac{1}{\sigma_v\sqrt{2\pi}} \exp\left\{- \frac{(L-D_f)^2}{2\sigma_v^2}\right\} \sum_{q=1}^{Q_j} \prod_{i \notin s_{qj}} p_i \prod_{i \in s_{qj}} (1-p_i) \frac{1}{RISK_2}$$

..(A4)

for all $j = 1, \ldots, n'$ and $L > G_j$

The expected deficiency is obtained by combining A4 and A1, giving

$$E\{u\} = \frac{1}{RISK_2} \sum_{j=1}^{n'} \left[\sum_{q=1}^{Q_j} \prod_{i \notin s_{qj}} p_i \prod_{i \in s_{qj}} (1-p_i) \right].$$

$$\int_{G_j}^{\infty} \frac{(L-G_j)}{\sigma_v\sqrt{2\pi}} \exp\left\{ - \frac{(L-D_f)^2}{2\sigma_v^2} \right\} dL \qquad (A5)$$

Now consider the integral

$$I = \int_{G_j}^{\infty} \frac{(L-G_j)}{\sigma_v\sqrt{2\pi}} \exp\left\{ - \frac{(L-D_f)^2}{2\sigma_v^2} \right\} dL$$

This may be rewritten as ..

$$I = \int_{G_j}^{\infty} \frac{(L-D_f)}{\sigma_v\sqrt{2\pi}} \exp\left\{-\frac{(L-D_f)^2}{2\sigma_v^2}\right\} dL - G_j \int_{G_j}^{\infty} \frac{1}{\sigma_v\sqrt{2\pi}} \exp\left\{-\frac{(L-D_f)^2}{2\sigma_v^2}\right\} dL$$

$$+ D_f \int_{G_j}^{\infty} \frac{1}{\sigma_v\sqrt{2\pi}} \exp\left\{-\frac{(L-D_f)^2}{2\sigma_v^2}\right\} dL \qquad \text{(A6)}$$

Substituting $x = (L-D_f)^2$ in the first term of equation A6 gives

$$I = \int_{(G_j-D_f)^2}^{\infty} \frac{1}{2\sigma_v\sqrt{2\pi}} \exp\left\{-\frac{x}{2\sigma_v^2}\right\} dx + (D_f-G_j) \int_{G_j}^{\infty} \frac{1}{\sigma_v\sqrt{2\pi}} \exp\left\{\frac{-(L-D_f)^2}{2\sigma_v^2}\right\} dL \qquad \text{....(A7)}$$

and equation A7 becomes, on solution of the integrals

$$I = \frac{\sigma_v}{\sqrt{2\pi}} \exp\left\{-\frac{(G_j-D_f)^2}{2\sigma_v^2}\right\} + \left\{\frac{D_f-G_j}{2}\right\}\left(1 - \text{erf}\left\{\frac{G_j-D_f}{\sigma_v\sqrt{2}}\right\}\right) \qquad \text{....(A8)}$$

Hence the expected deficiency becomes

$$E\{u\} = \frac{1}{RISK_2}\sum_{j=1}^{n'} \left(\sum_{q=1}^{Q_j} \prod_{i \notin s_{qj}} p_i \prod_{i \in s_{qj}} (1 - p_i) \right) .$$

$$.\left\{\frac{D_f-G_j}{2}\right\}\left(1-\text{erf}\left\{\frac{G_j-D_f}{\sigma_v\sqrt{2}}\right\} + \frac{\sigma_v}{\sqrt{2\pi}}\exp\left\{-\frac{(G_j-D_f)^2}{2\sigma_v^2}\right\}\right) \qquad \text{....(A9)}$$

APPENDIX B

DERIVATION OF MEAN SQUARE ERROR OF THE KARHUNEN LOEVE EXPANSION

Given that

$$D_{jt} = \sum_{k=1}^{K} a_{jk}\lambda_k^{\frac{1}{2}} \phi_{kt} + \mathcal{E}_{jt} \qquad t = 1, \ldots, M \qquad \text{(B1)}$$

Then the mean square error, E, averaged over the N samples and M intervals is

$$E = \frac{1}{N.M} \sum_{j=1}^{N} \sum_{t=1}^{M} \left(D_{jt} - \sum_{k=1}^{K} a_{jk}\lambda_k^{\frac{1}{2}} \phi_{kt} \right)^2 \qquad \text{(B2)}$$

This error is a minimum when the first order variations with respect to $(a_{jk} . \lambda_k^{\frac{1}{2}})$ and ϕ_{kt} are zero, i.e. when

$$\frac{\partial E}{\partial a_{jl}\lambda_l^{\frac{1}{2}}} = \frac{1}{N.M} \sum_j \sum_t \left(D_{jt} - \sum_k a_{jk} \lambda_k^{\frac{1}{2}} \phi_{kt}\right) -2\phi_{lt}$$

$$= 0 \qquad \text{for } l = 1, \ldots, K$$

$$\text{Therefore} \quad \sum_j \sum_t D_{jt} \phi_{lt} = \sum_j \sum_t \phi_{lt} \sum_k a_{jk}\lambda_k^{\frac{1}{2}} \phi_{kt} \qquad l=1,\ldots,K \qquad \text{(B3)}$$

and when

$$\frac{\partial E}{\partial \phi_{lt}} = \frac{1}{N M} \sum_j \sum_t \left(D_{jt} - \sum_k a_{jk} \lambda_k^{\frac{1}{2}} \phi_{kt}\right) -2a_{jl}\lambda_l^{\frac{1}{2}} \qquad l=1, \ldots, K$$

$$= 0$$

Therefore

$$\sum_j \sum_t D_{jt}\, a_{jl}\, \lambda_l^{\frac{1}{2}} = \sum_j \sum_t a_{jl}\, \lambda_l^{\frac{1}{2}} \quad \sum_k a_{jk}\, \lambda_k^{\frac{1}{2}}\, \phi_{kt}$$

$$\text{for } l = 1, \ldots, K \qquad \text{(B4)}$$

Assume the vectors $\underset{\sim}{\phi_k} = \left\{\phi_{k1}, \phi_{k2} \ldots \phi_{kM}\right\}^T$ are orthonormal.

Hence

$$\sum_{s,t=1}^{M} \phi_{kt} \cdot \phi_{ks} \quad \begin{aligned} &= 1 \quad \text{for } t = s \\ &= 0 \quad \text{for } t \neq s \end{aligned} \qquad \text{(B5)}$$

and assume the coefficients a_{jk} to be uncorrelated over the samples, i.e.

$$\sum_{j=1}^{N} a_{jk}\, a_{jl} \quad \begin{aligned} &= N \quad \text{for } k = l \\ &= 0 \quad \text{for } k \neq l \end{aligned} \qquad \text{(B6)}$$

Equation B3 then becomes

$$\sum_j \sum_t D_{jt}\, \phi_{lt} = \sum_j \lambda_l^{\frac{1}{2}}\, a_{jl}$$

Therefore

$$a_{jl} = \frac{1}{\lambda_l^{\frac{1}{2}}} \sum_t D_{jt}\, \phi_{lt} \quad \text{for } j = 1, \ldots, N \qquad \text{(B7)}$$

$$l = 1, \ldots, K$$

and equation B4 becomes

$$\sum_j \sum_t D_{jt}\, a_{jl}\, \lambda_l^{\frac{1}{2}} = \sum_t N \cdot \lambda_l \cdot \phi_{lt}$$

Therefore

$$\lambda_l^{\frac{1}{2}}\, \phi_{lt} = \frac{1}{N} \sum_j D_{jt}\, a_{jl} \quad \text{for } t = 1, \ldots, M$$

$$l = 1, \ldots, K \qquad \text{(B8)}$$

Eliminating a_{jl} from B7 and B8 gives

$$\lambda_l^{\frac{1}{2}} \phi_{lt} = \frac{1}{N} \sum_j D_{jt} \left(\frac{1}{\lambda_l^{\frac{1}{2}}} \sum_s D_{js} \phi_{ls} \right)$$

Therefore

$$\lambda_l \phi_{lt} = \sum_s \left(\frac{1}{N} \sum_j D_{jt} D_{js} \right) \phi_{ls} \qquad \text{(B9)}$$

After writing $R_{ts} = \frac{1}{N} \sum_j D_{jt} D_{js} \quad t,s = 1, \ldots, M$ (B10)

equation B9 becomes

$$\lambda_l \phi_{lt} = \sum_s R_{ts} \phi_{ls} \qquad \text{for } s,t = 1, \ldots, M \qquad \text{(B11)}$$
$$l = 1, \ldots, K$$

Equation B11 is an eigenvalue equation, the λ_l and ϕ_{ls} being the eigenvalues and vectors respectively of the covariance matrix R_{ts}. Combining equations B2 and B11 gives

$$E = \frac{1}{M} \sum_t \left[\frac{1}{N} \sum_j \left\{ D_{jt}^2 - 2D_{jt} \sum_k a_{jk} \lambda_k^{\frac{1}{2}} \phi_{kt} + \left(\sum_k a_{jk} \lambda_k^{\frac{1}{2}} \phi_{kt} \right)^2 \right\} \right]$$

$$= \frac{1}{M} \left[\sum_t R_{tt} - 2 \sum_k \sum_t \phi_{kt} \sum_s \frac{1}{N} \sum_j D_{jt} D_{js} \phi_{ks} + \sum_k \lambda_k \right]$$

$$= \frac{1}{M} \left[\sum_t R_{tt} - 2 \sum_k \lambda_k \sum_t \phi_{kt}^2 + \sum_k \lambda_k \right]$$

$$= \frac{1}{M} \left[\sum_t R_{tt} - \sum_k \lambda_k \right] \qquad \text{(B12)}$$

Equation B12 gives the mean square error of the expansion over the samples and time which is introduced by truncating the series at the Kth term.

APPENDIX C

RESULTS OF EXPERIMENTAL FORECASTS MADE USING FARMER'S METHOD

Any member of an ensemble of N sample load curves may be represented by equation B1 (see Appendix B) to an accuracy given by equation B12. The forecasting problem is now to predict further load curves, which are assumed to be generated by the same process as the sample functions. This requires that the a_{jt} (equation B7) be determined for $j > N$. The approach used in the experiments was to find these so that equation B2 was satisfied in a least squares sense over the L most recent demand intervals (72, 75). As the daily load curves are contiguous the past history L can span from one day to the next. Thus if present time is denoted by M_o the problem is to minimize, by choice of $\hat{a}_{jk}$,

$$Q = \sum_{t=M_o-L}^{M_o} \left[D_{jt} - \sum_{k=1} \hat{a}_{jk} \lambda_k^{\frac{1}{2}} \phi_{kt} \right] \qquad \text{(C1)}$$

for $j > N$

The demand can then be predicted from

$$\hat{D}_{jt} = \sum_{k=1}^{K} \hat{a}_{jk} \lambda_k^{\frac{1}{2}} \phi_{kt} \qquad t = M_o + 1, \ldots, M_o + r \qquad \text{(C2)}$$

where the lead time, r, may be as long as desired, although practically it should not exceed 24 hours.

The experimental forecasts used half hourly integrated demands for the Christchurch M.E.D. and Riccarton Borough Council distribution systems. Half-hourly forecasts were made for 10 days from 25.5.67. The eigenvectors and values were calculated from the 10, 20 or 30 days preceding this date. Weekends and holidays were excluded as these days have different load patterns from ordinary weekdays.

The effect on the prediction error, given by

$$e_p = 100 \times (D_{jt} - \hat{D}_{jt}) / D_{jt} \qquad (C3)$$

of varying N, K and L was investigated for the two distribution systems for lead times of 1, 5 and 10 hours. The distribution of e_p for K = 1 and 2, N = 10 and L = 10 and 20 is shown in figures C1 and C2 while a summary of the mean e_p and standard deviation $\sigma(e_p)$ of this distribution for all values of K, L and N studied is given in tables C1 and C2.

Each of the N sample load curves can be considered as a vector $\underset{\sim}{D}_j$ in M-dimensional space. The quantity $\lambda_k^{\frac{1}{2}}$ is the RMS length of the projections of the $\underset{\sim}{D}_j$ onto $\underset{\sim}{\phi}_k$. The eigen-vector associated with the largest eigenvalue represents the most significant feature of the process; the remaining eigen-vectors represent progressively less significant features. The maximum number of non-zero eigenvalues of R (the covariance matrix) is N, for $N < M$, or M if $N \geqslant M$.

If the shape (but not necessarily the magnitude) of the daily load curve is unchanged over the sample set then all the $\underset{\sim}{D}_j$, $j=1, \ldots, N$, have the same direction. In this case only the first λ_k is non-zero and the process can be represented by

$$D_{jt} = a_{j1} \quad \lambda_1^{\frac{1}{2}} \phi_{1t} \tag{C4}$$

The labour involved in solving for the eigenvectors is not justified as

$$\lambda_1^{\frac{1}{2}} \cdot \phi_{1t} = \frac{1}{N} \sum_{j=1}^{N} D_{jt} \qquad t = 1, \ldots, M \tag{C5}$$

The shape of the load curve is not constant from day to day due to random weather fluctuations and seasonal changes in the environment (e.g. changes in the duration of daylight). The $\underset{\sim}{D}_j$, $j = 1, \ldots, N$ thus have differing directions. The dominant term in the expansion can still be obtained using equation C5 while the remaining terms represent deviations from this mean. The relative contribution of each term is determined by the ratio

$$(\lambda_k / \lambda_1)^{\frac{1}{2}} \quad , \quad k > 1$$

This is shown in table C3 for the values of N.

The assumption involved in finding the $\hat{a}_{jk}$, $j > N$ via equation C1 means that, strictly, the method can only be used on future days when the environment is similar to that on the sample days. This is unlikely and tables C1 and C2 indicate that the reduction in $\sigma(e_p)$ obtained by increasing K is marginal. Generally the λ_k, $\underset{\sim}{\phi}_k$ for $k > 1$ are specific to the conditions which produced the sample load curves and do not add to the knowledge of the future.

Increasing L has the effect of "smoothing" random fluctuations in the demand record. This is advantageous when these may be large but of zero mean. In a situation where the environment is changing rapidly a small value of L may be used to increase the rate of response to change.

For lead times of 5 or 10 hours $\sigma(e_p)$ increases with K for constant L and N. The rate of increase is reduced if N is also increased. Table C4 illustrates this for selected values of N and L.

There is no significant gain in forecasting accuracy from using K > 1 and a considerable amount of computation may be avoided by calculating $\lambda_1^{\frac{1}{2}}$, ϕ_{1t} from equation C5. The method of spectral expansion is then equivalent to the scaling technique used by Farmer (73) to overcome problems with data transmission errors. This technique scales a standard load curve with the ratio of the actual demand on the day of prediction to the corresponding value of the standard curve. This is equivalent to obtaining $\hat{a}_{j1}$ for $j > N$ from

$$\hat{a}_{j1} = \frac{D_{jM_o}}{\lambda_1^{\frac{1}{2}} \phi_{1M_o}} \qquad j > N \ , \ t = M_o \qquad \text{(C6)}$$

Forecasts are then obtained for $t > M_o$ from;

$$\hat{D}_{jt} = \hat{a}_{j1} \lambda_1^{\frac{1}{2}} \phi_{1t} \qquad t > M_o \ , \ j > N \qquad \text{(C7)}$$

Sample Curves N	Historical Data L	Number of modes K: 1	2	4	8
10	10	- 0.3 6.2	- 0.2 6.6	0.8 5.5	1.8 10.2
	20	- 0.2 6.4	- 0.1 6.1	1.3 6.9	1.5 6.1
	30	- 0.1 6.5	0.3 6.1	0.4 6.0	0.9 5.6
20	10	0.2 6.4	- 0.2 6.9	0.2 5.9	0.1 11.6
	20	0.2 6.6	0.0 6.7	0.8 8.0	0.5 5.7
	30	- 0.3 6.7	1.0 6.7	- 0.3 7.7	0.3 6.0
30	10	- 0.3 6.8	0.2 7.2	0.2 5.9	0.2 12.0
	20	- 0.3 7.1	0.4 7.3	0.7 7.6	- 0.4 7.3
	30	- 0.4 7.2	1.4 7.6	0.1 6.9	- 0.4 7.3

TABLE C1. Mean and standard deviation of percentage forecasting error over 480 forecasts. Lead time 1 hour, Christchurch M.E.D. demand data.

Sample curves N	Historical data L	Number of Modes K 1	2	4	8
10	10	- 0.2, 5.8	- 0.8, 6.6	- 0.7, 7.5	- 0.3, 11.2
	20	- 0.2, 5.9	- 0.6, 5.7	- 0.0, 6.5	0.2, 6.3
	30	- 0.1, 6.2	- 0.4, 5.7	- 0.2, 6.4	0.6, 6.3
20	10	- 0.2, 6.1	- 0.7, 6.0	- 0.4, 8.0	- 0.1, 14.2
	20	- 0.2, 6.3	0.0, 6.0	0.2, 6.8	0.6, 6.4
	30	- 0.3, 6.3	0.1, 6.0	0.6, 6.8	0.4, 6.4
30	10	- 0.3, 6.4	- 0.2, 6.2	- 0.0, 8.0	- 0.0 12.9
	20	- 0.2, 6.3	0.3, 6.4	0.6, 7.5	0.3, 7.8
	30	- 0.5, 6.6	0.5, 6.5	1.1, 7.2	0.3, 6.6

Table C2 Mean and standard deviation of percentage forecasting error over 480 forecasts. Lead time 1 hour. Riccarton B.C. demand data.

	N = 10	N = 20	N = 30
K			
1	1.0	1.0	1.0
2	3.6×10^{-3}	3.01×10^{-3}	2.61×10^{-3}
3	1.5×10^{-3}	1.12×10^{-3}	0.95×10^{-3}
4	2.63×10^{-4}	0.92×10^{-3}	0.91×10^{-3}
5	1.53×10^{-4}	1.88×10^{-4}	3.78×10^{-4}
6	0.82×10^{-4}	1.06×10^{-4}	1.62×10^{-4}
7	2.84×10^{-5}	5.76×10^{-5}	0.74×10^{-4}
8	1.96×10^{-5}	5.54×10^{-5}	6.0×10^{-5}
10	0.94×10^{-5}	1.83×10^{-5}	2.46×10^{-5}
12		0.97×10^{-5}	1.88×10^{-5}
16		2.96×10^{-6}	6.85×10^{-6}

Table C3 Ratio $(\lambda_k / \lambda_1)^{\frac{1}{2}}$
for Christchurch M.E.D. data

K = 1	L = 10	L = 20	L = 30
N = 10	7.9	7.6	7.2
N = 20	8.2	7.9	7.4
N = 30	8.7	8.5	7.9

(a) Lead time is 5 hours.

K = 1	L = 10	L = 20	L = 30
N = 10	8.5	7.6	6.9
N = 20	8.7	7.8	7.1
N = 30	9.4	8.5	7.4

(b) Lead time is 10 hours

Table C4. Variation of $\sigma(e_p)$ with lead time.

Christchurch M.E.D. demand data.

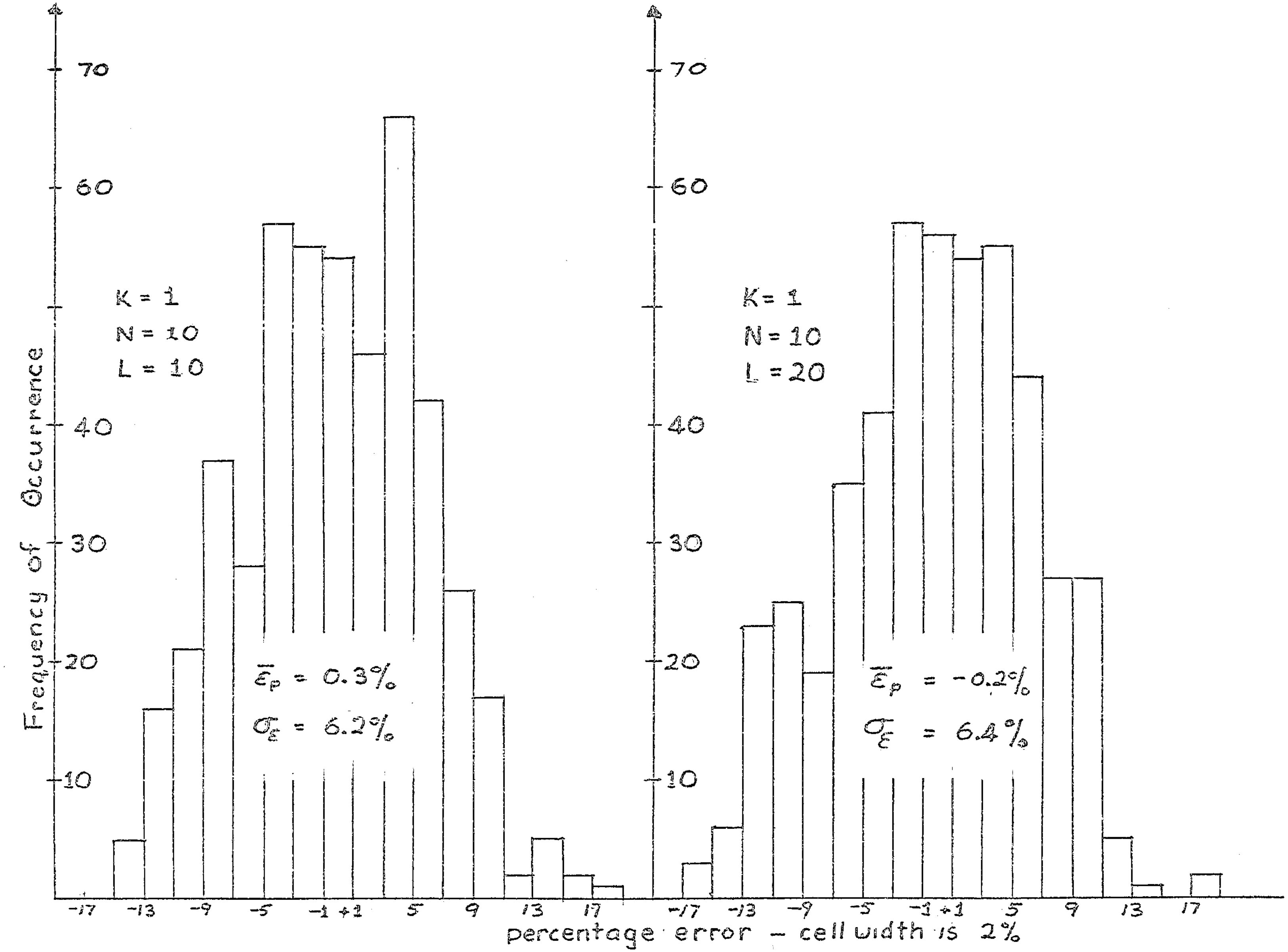


Figure C1: Christchurch MED data, lead time 1 hour.

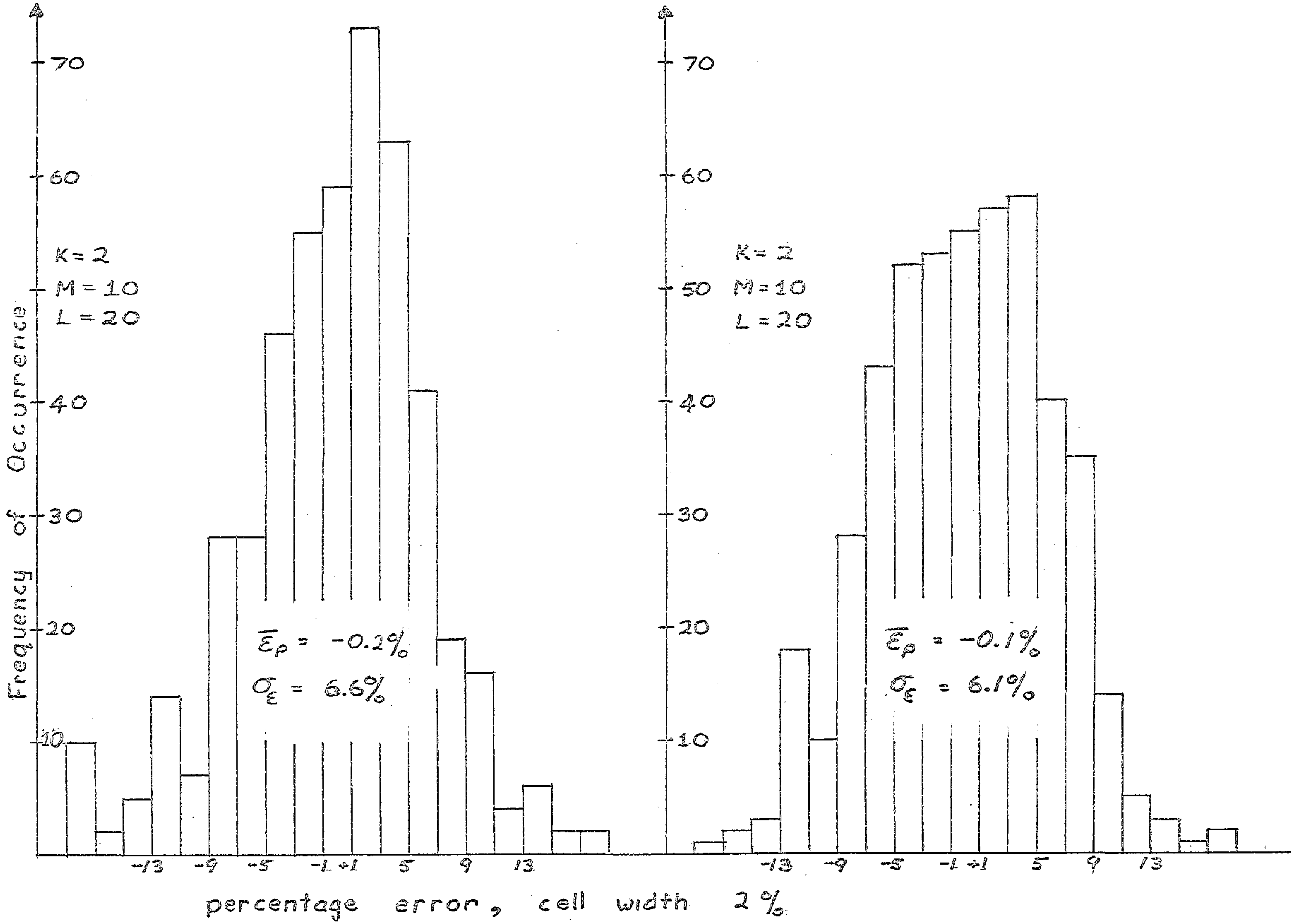


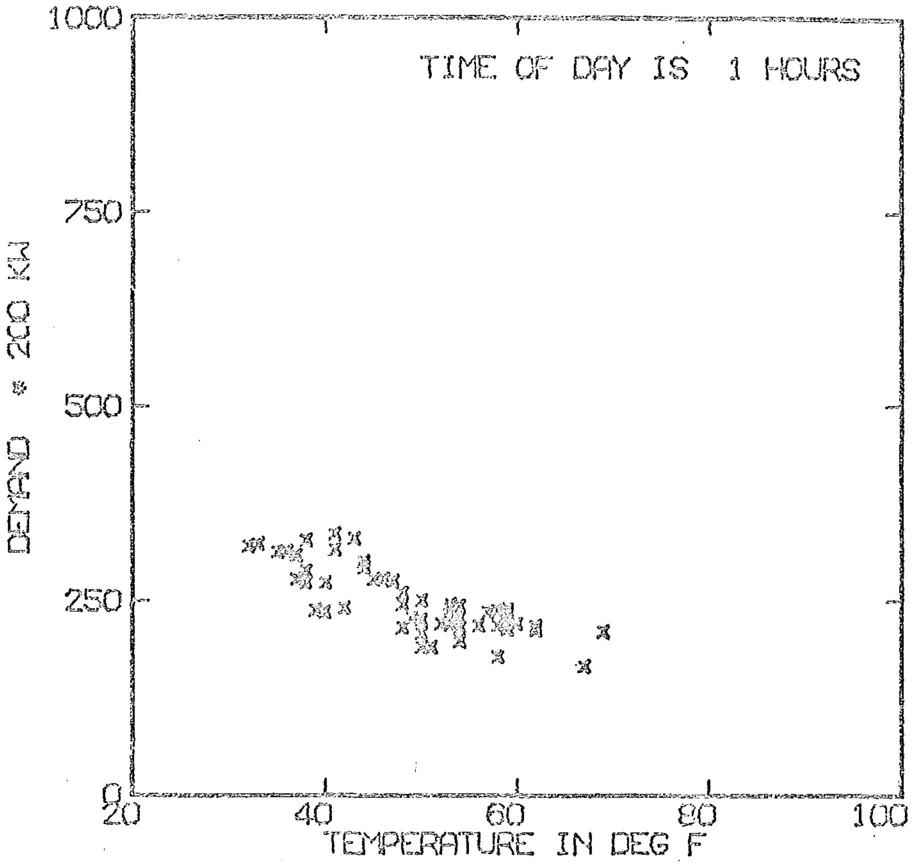
Figure C2: Christchurch MED data, lead time 1 hour.

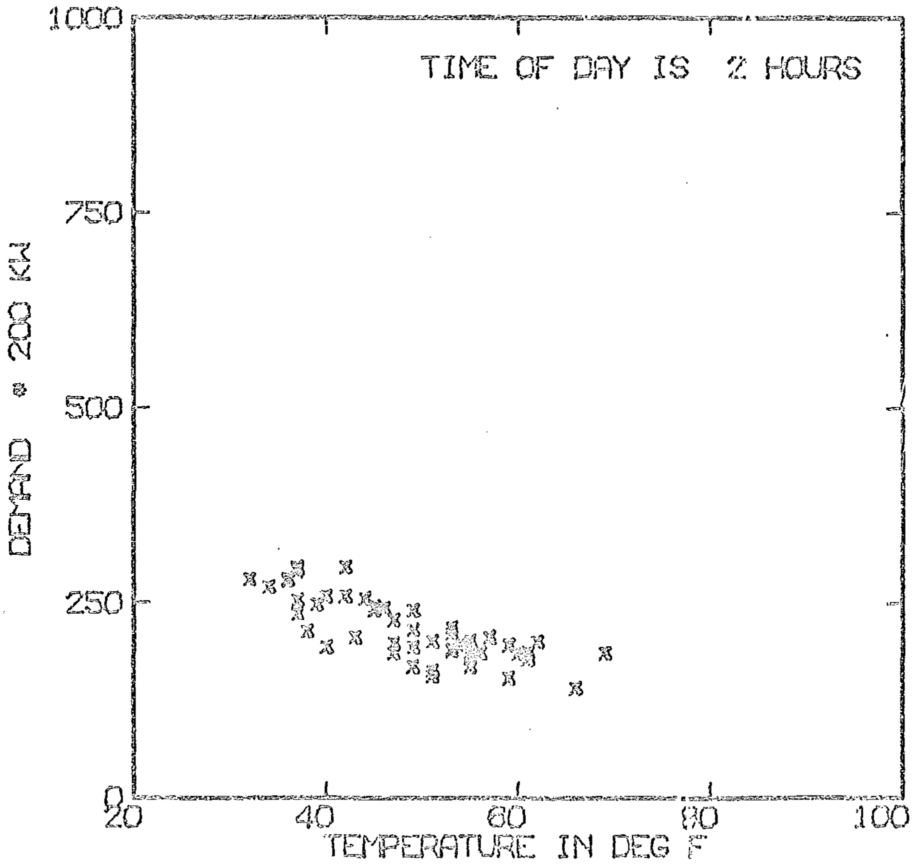
APPENDIX D

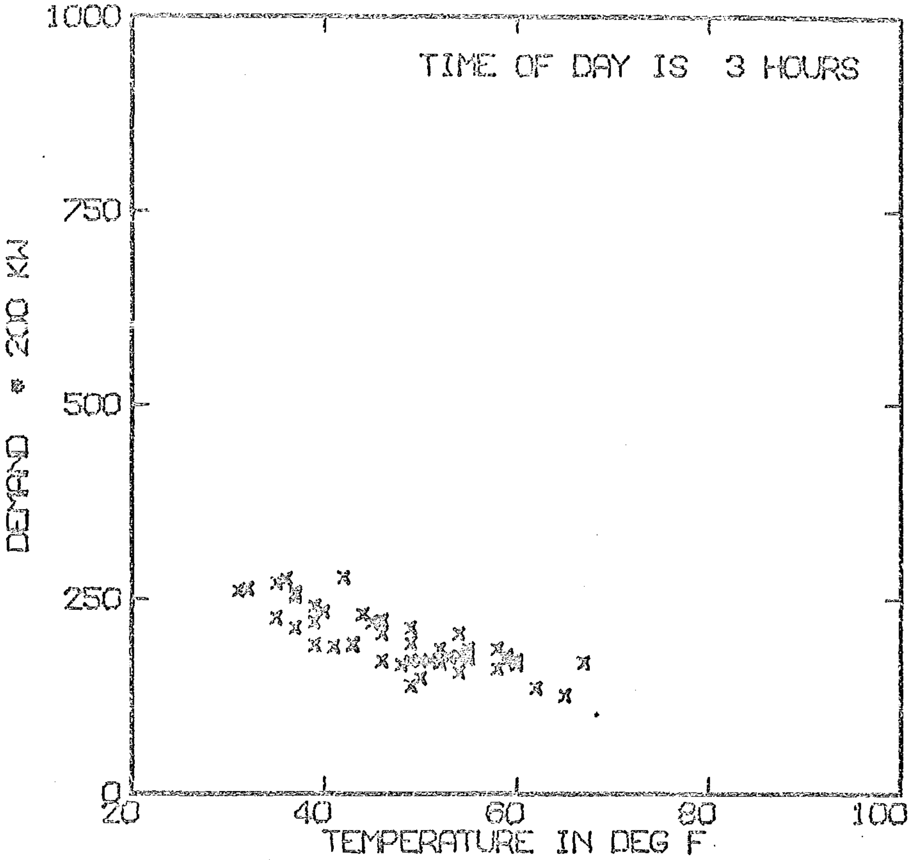
DEMAND - TEMPERATURE SCATTER DIAGRAMS FOR AN URBAN LOAD

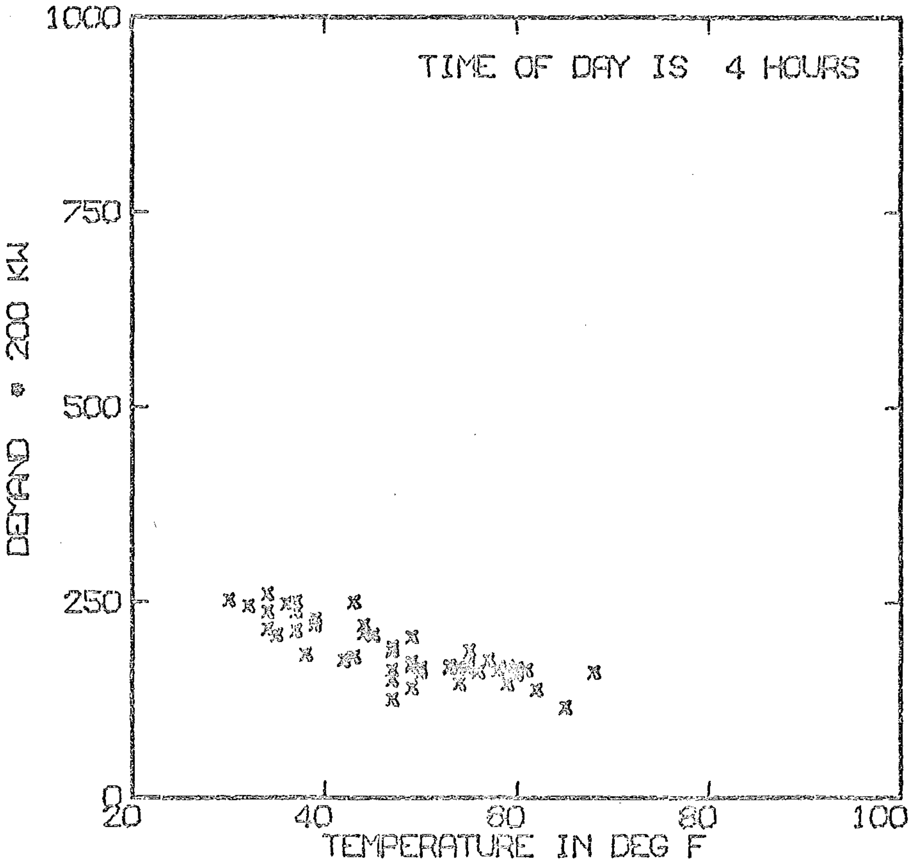
This appendix contains plots of the half hourly integrated demand, for the second half hour of each hourly period, against the ambient air temperature measured on the hour, over a 24 hour period. The load was the Christchurch M.E.D. The demands were those on 53 successive Tuesdays starting from 28 March 1967.

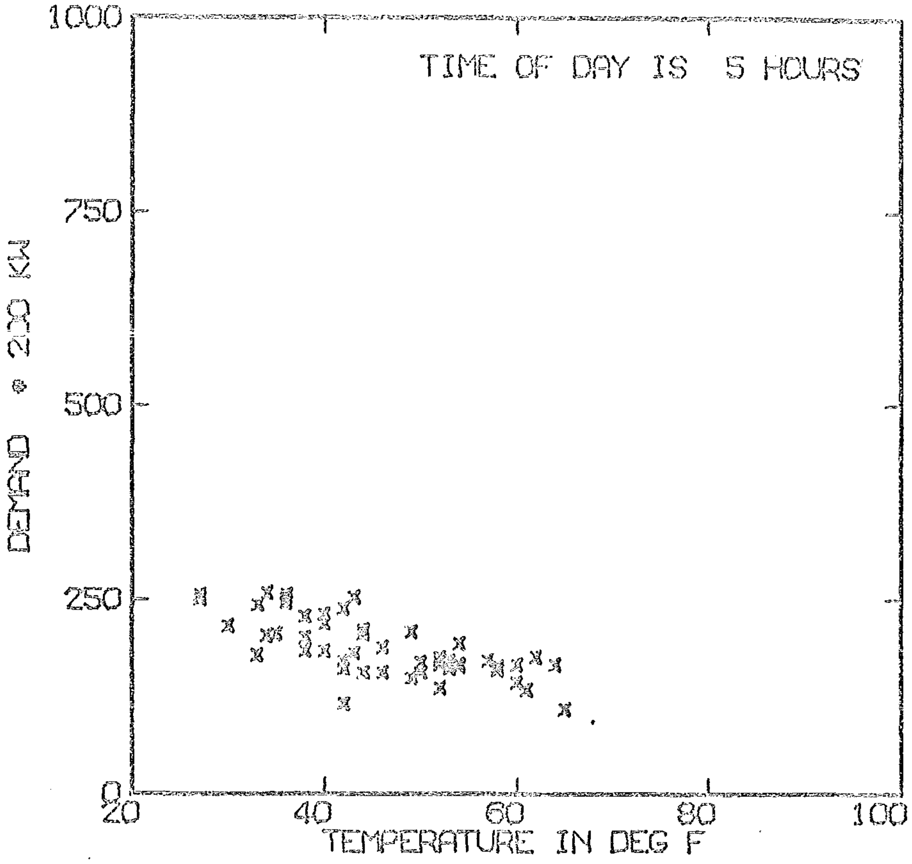
The demand - temperature pairs which occurred on 'day 28' (3.10.67) are identified on the plots for hours 0800 to 1700.

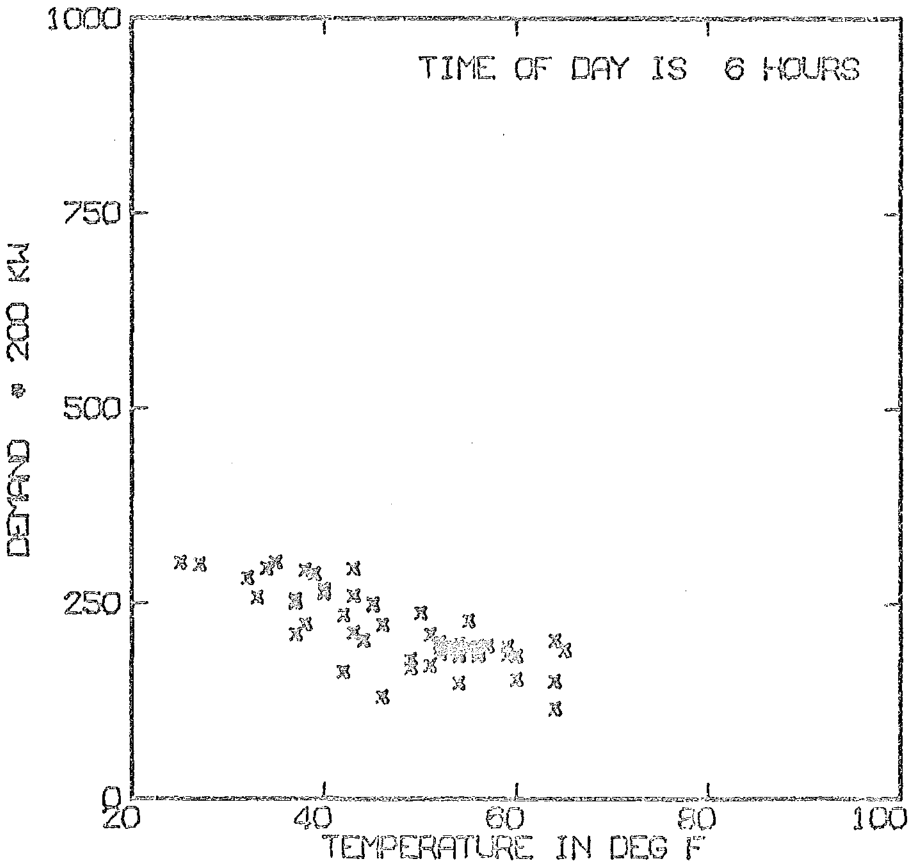

TIME OF DAY IS 1 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
0
250
500
750
1000
20
40
60
80
100

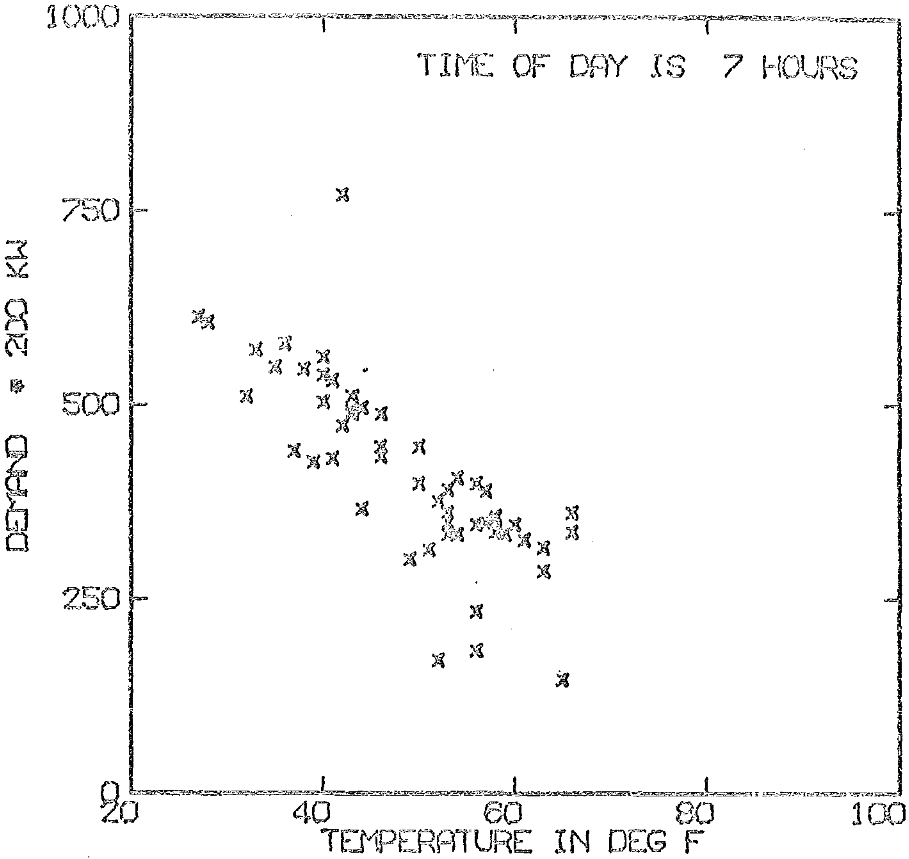

TIME OF DAY IS 2 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
0
250
500
750
1000
20
40
60
80
100


TIME OF DAY IS 3 HOURS
DEMAND * 200 KW
1000
750
500
250
0
20
40
60
80
100
TEMPERATURE IN DEG F

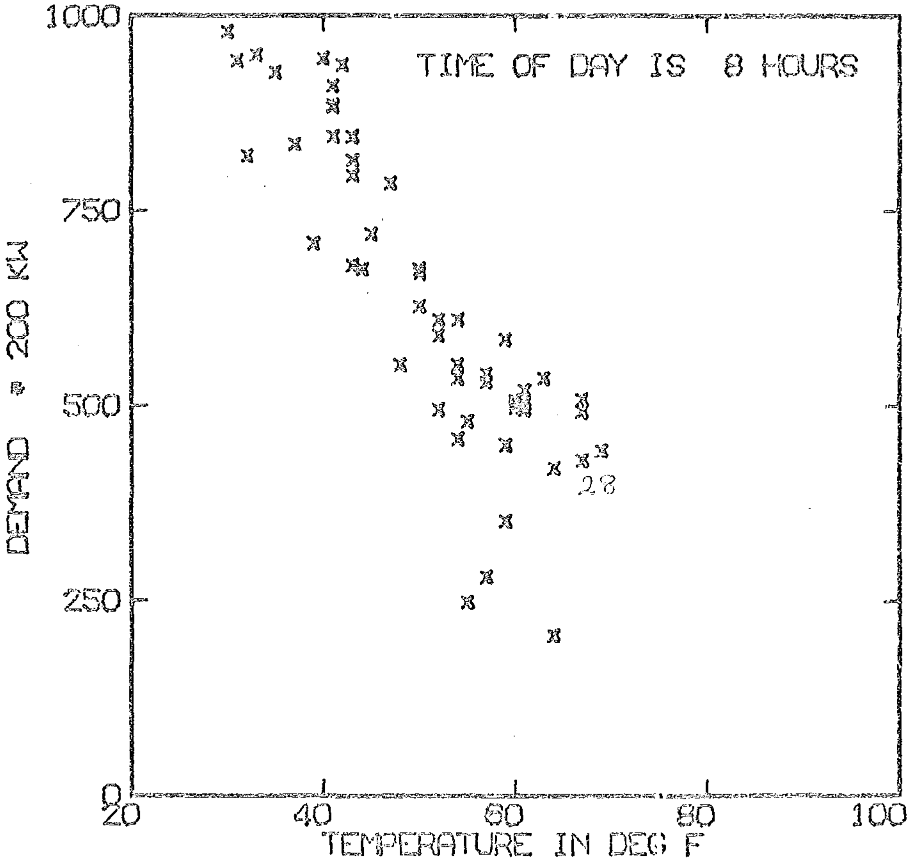

TIME OF DAY IS 4 HOURS
DEMAND * 200 KW
1000
750
500
250
0
20
40
60
80
100
TEMPERATURE IN DEG F


TIME OF DAY IS 5 HOURS
1000
750
500
250
0
DEMAND * 200 KW
20
40
60
80
100
TEMPERATURE IN DEG F

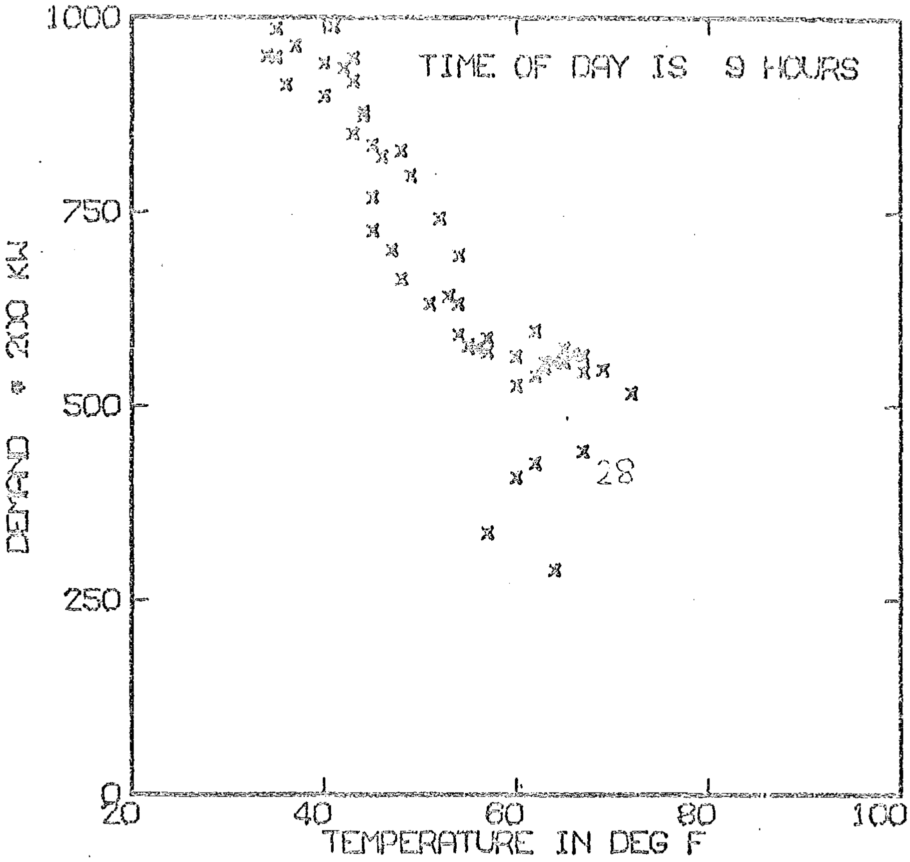

TIME OF DAY IS 6 HOURS
1000
750
500
250
0
DEMAND * 200 KW
20
40
60
80
100
TEMPERATURE IN DEG F

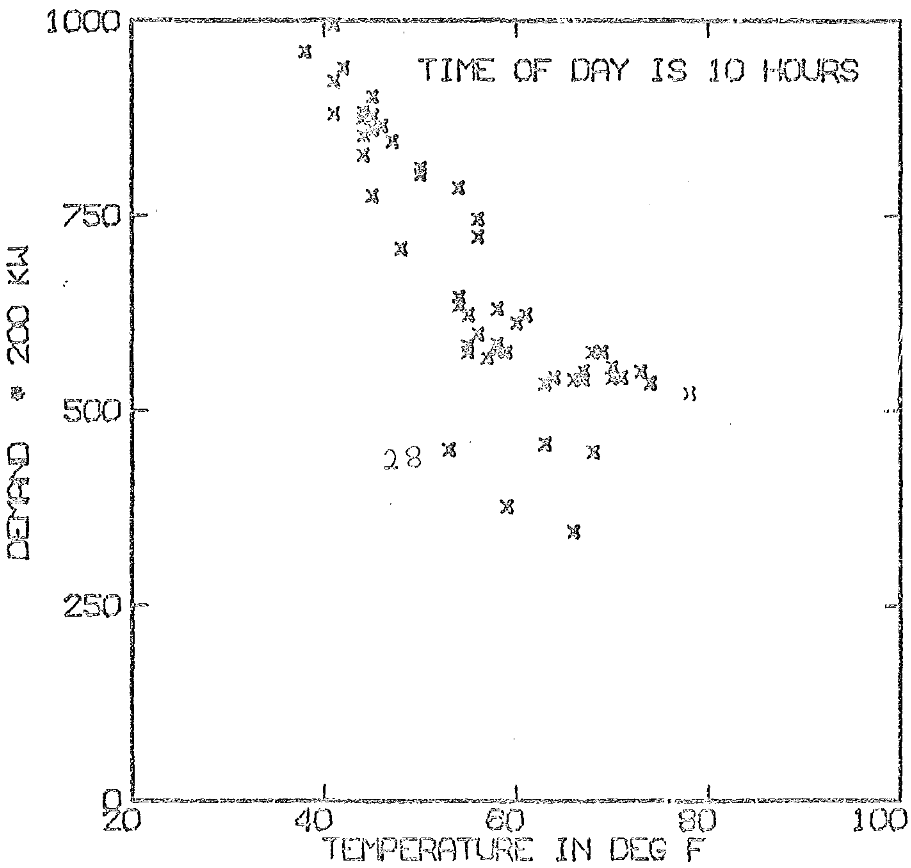


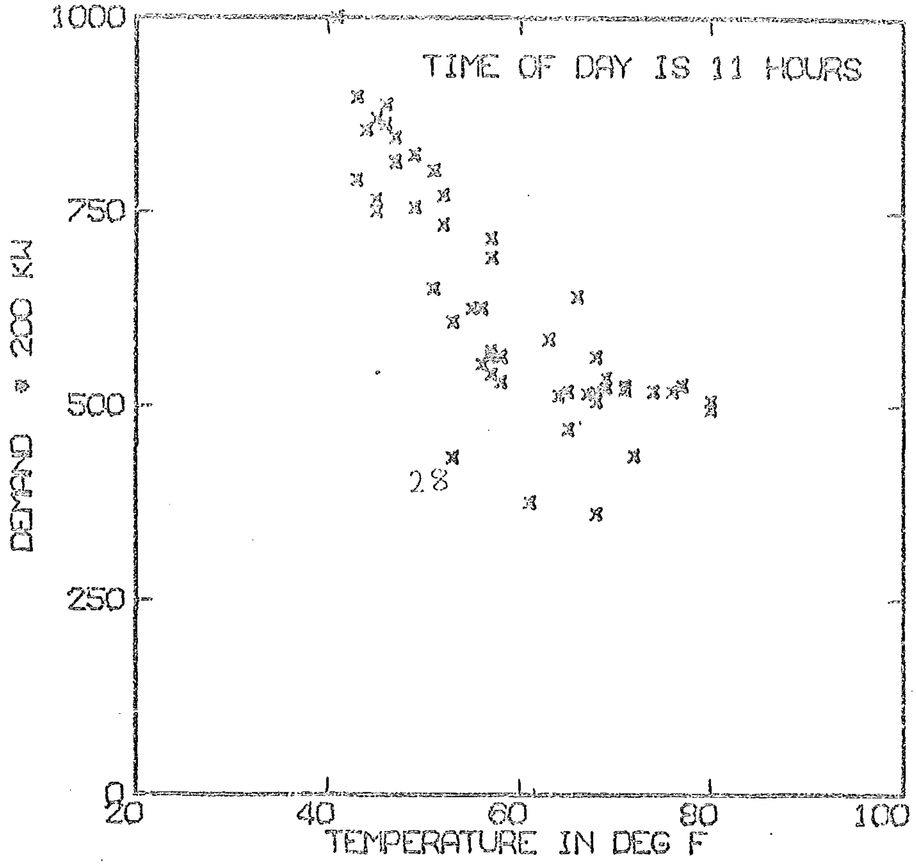
TIME OF DAY IS 7 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
1000
750
500
250
0
20
40
60
80
100

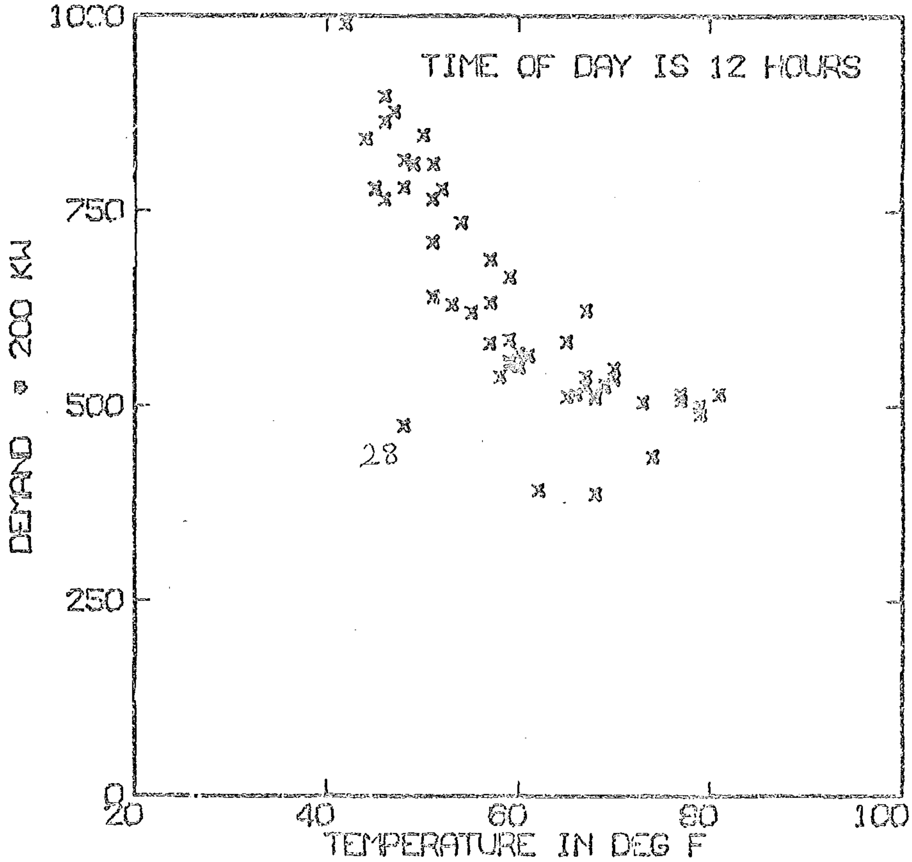


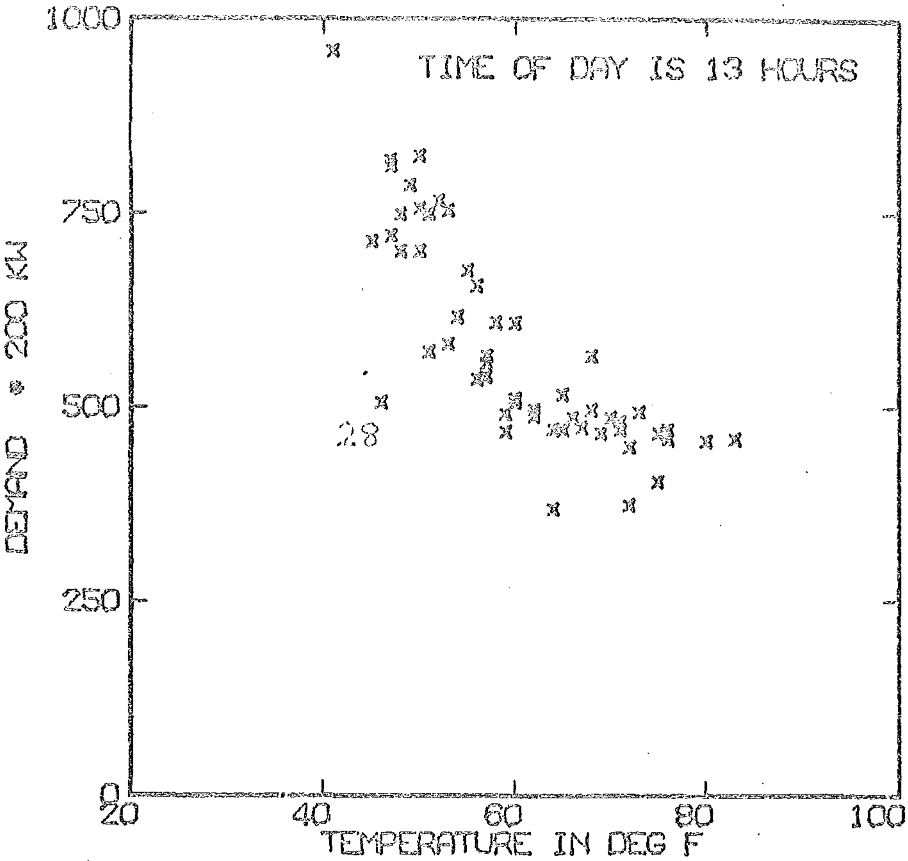
TIME OF DAY IS 8 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
28
1000
750
500
250
0
20
40
60
80
100

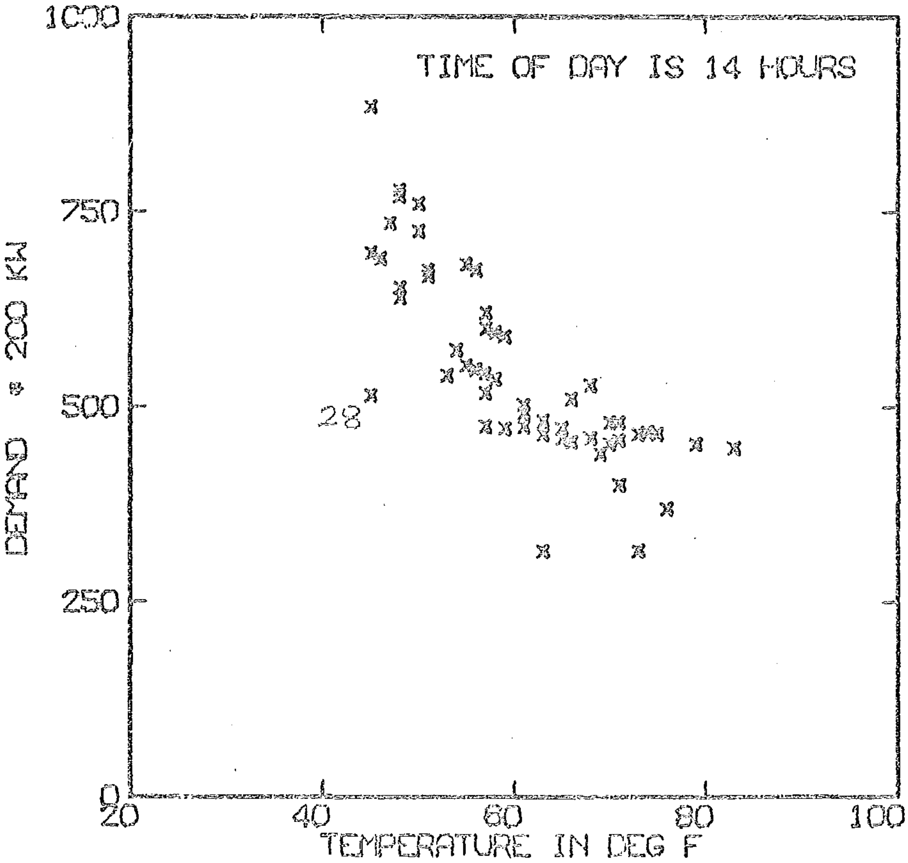

TIME OF DAY IS 9 HOURS
DEMAND * 200 KW
1000
750
500
250
0
20
40
60
80
100
TEMPERATURE IN DEG F
28

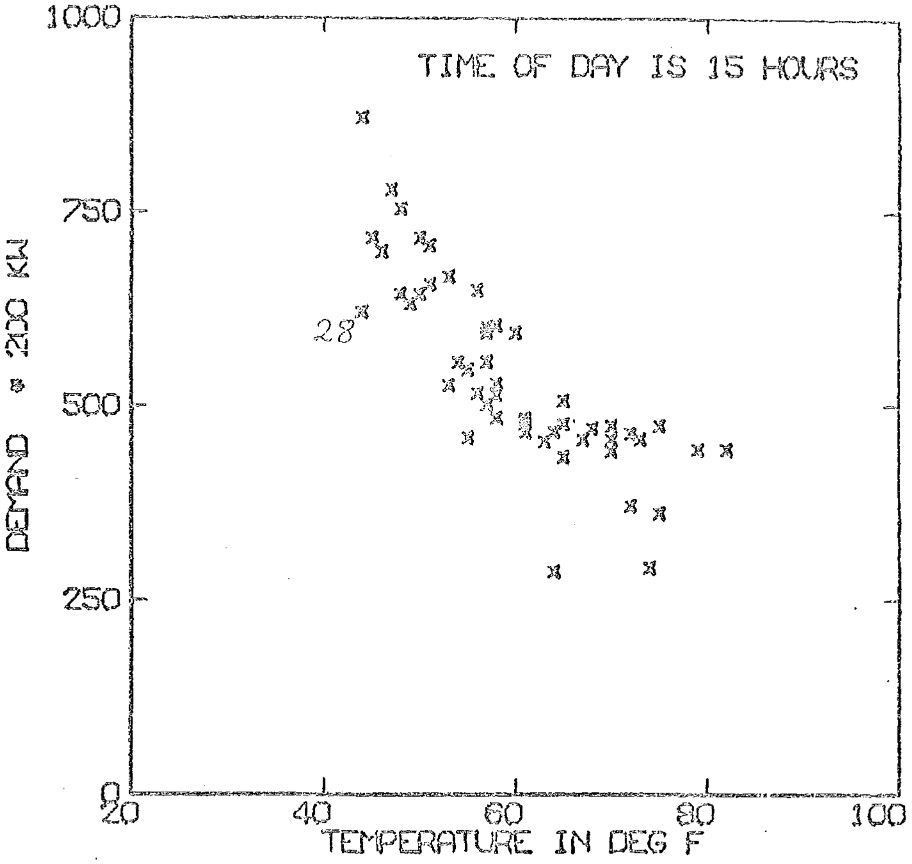

TIME OF DAY IS 10 HOURS
DEMAND * 200 KW
1000
750
500
250
0
20
40
60
80
100
TEMPERATURE IN DEG F
28

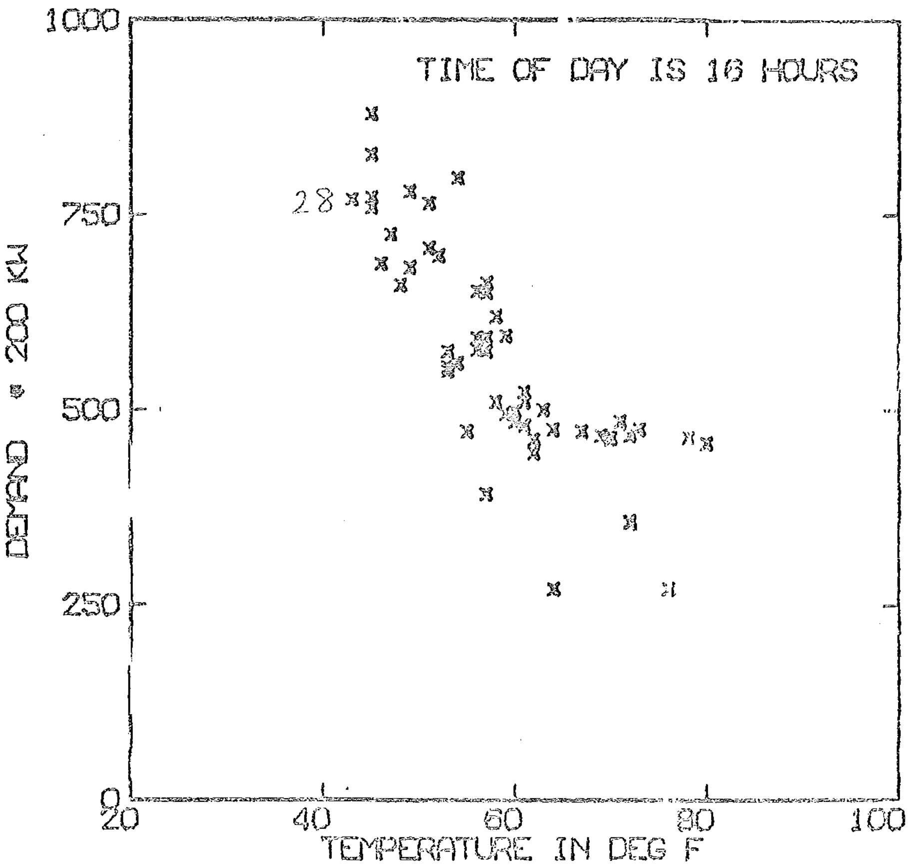

TIME OF DAY IS 11 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
1000
750
500
250
0
20
40
60
80
100
28

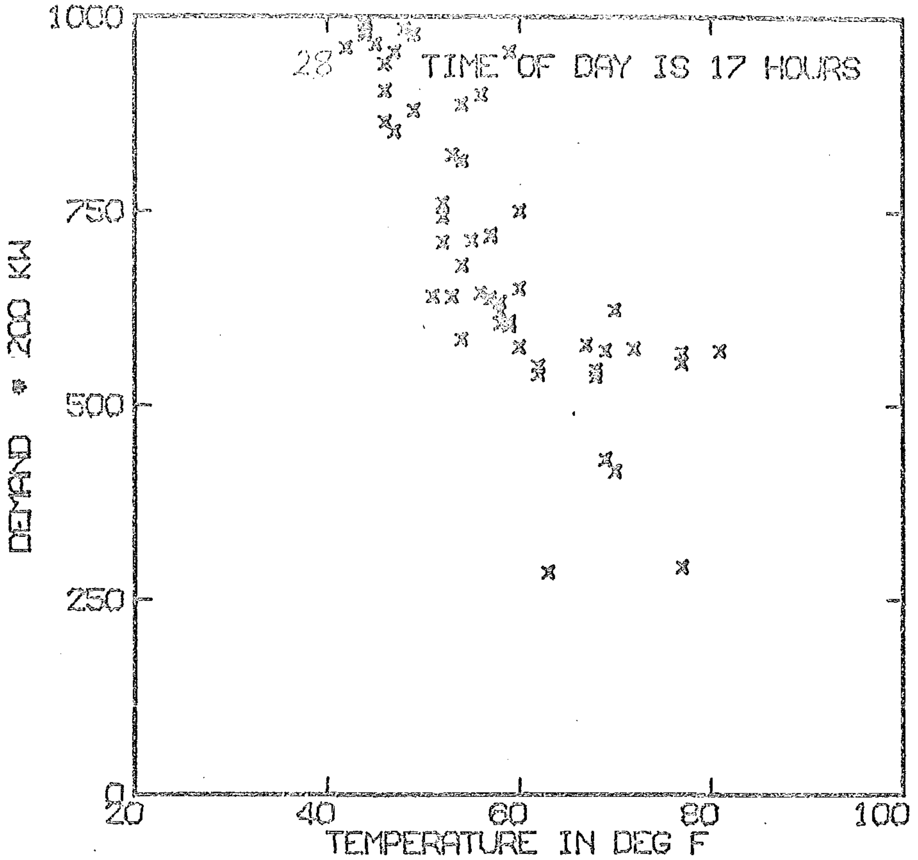

TIME OF DAY IS 12 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
1000
750
500
250
0
20
40
60
80
100
28

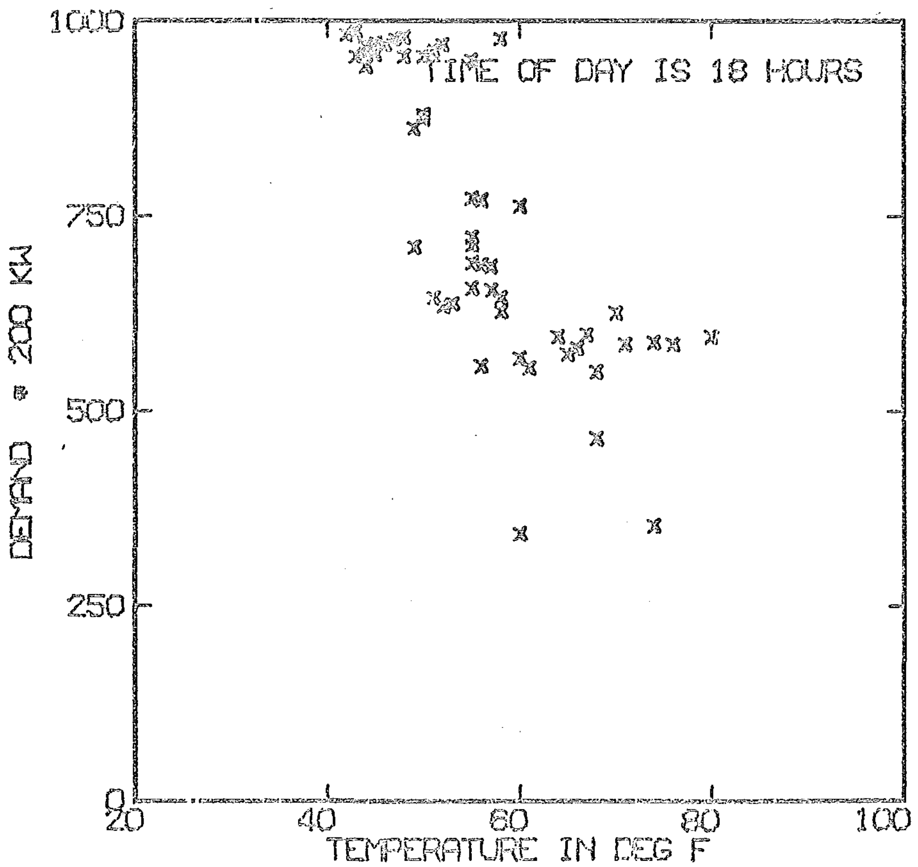

TIME OF DAY IS 13 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
28
0
250
500
750
1000
20
40
60
80
100

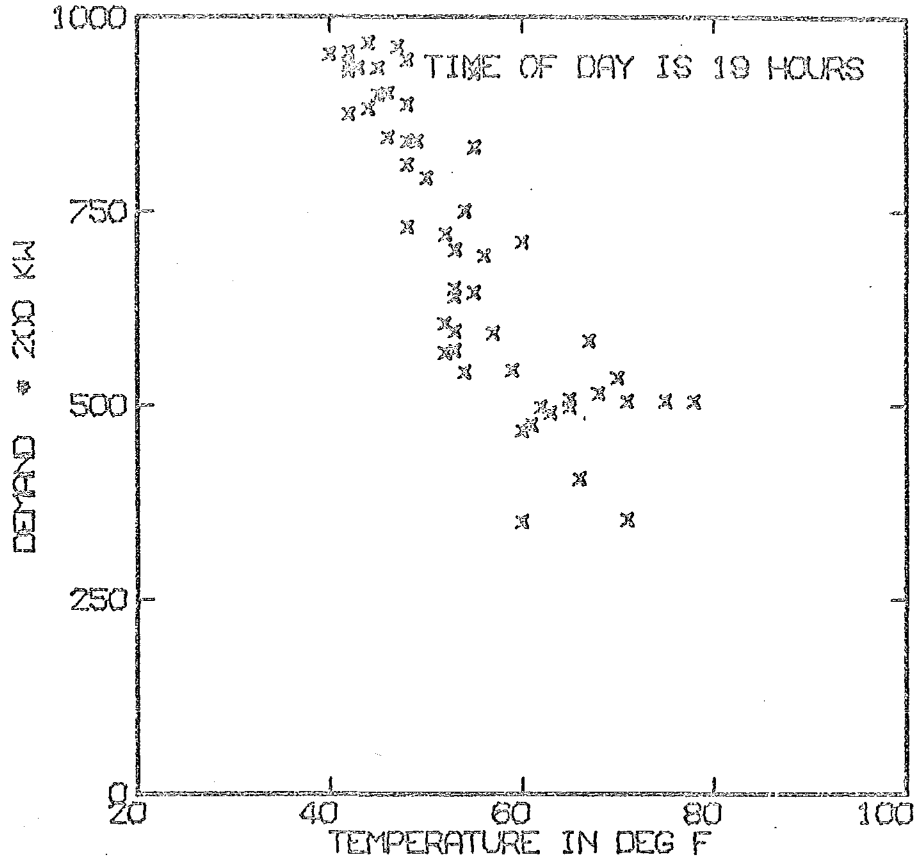

TIME OF DAY IS 14 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
28
0
250
500
750
1000
20
40
60
80
100

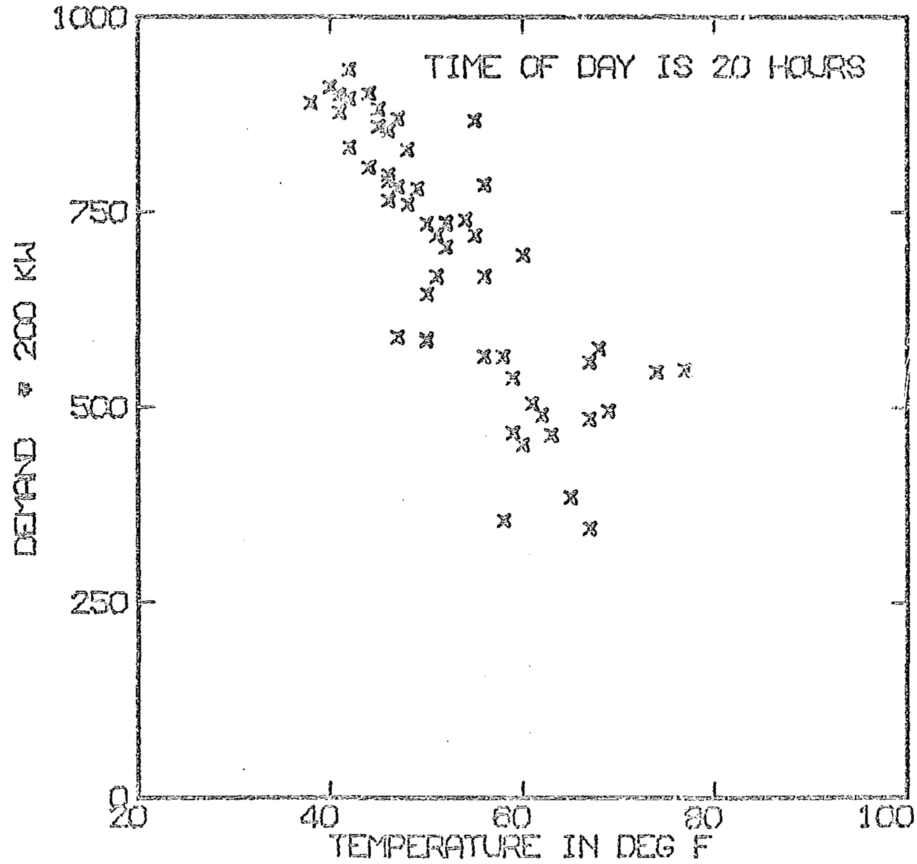

TIME OF DAY IS 15 HOURS
1000
750
500
250
0
20
40
60
80
100
DEMAND * 200 KW
TEMPERATURE IN DEG F
28

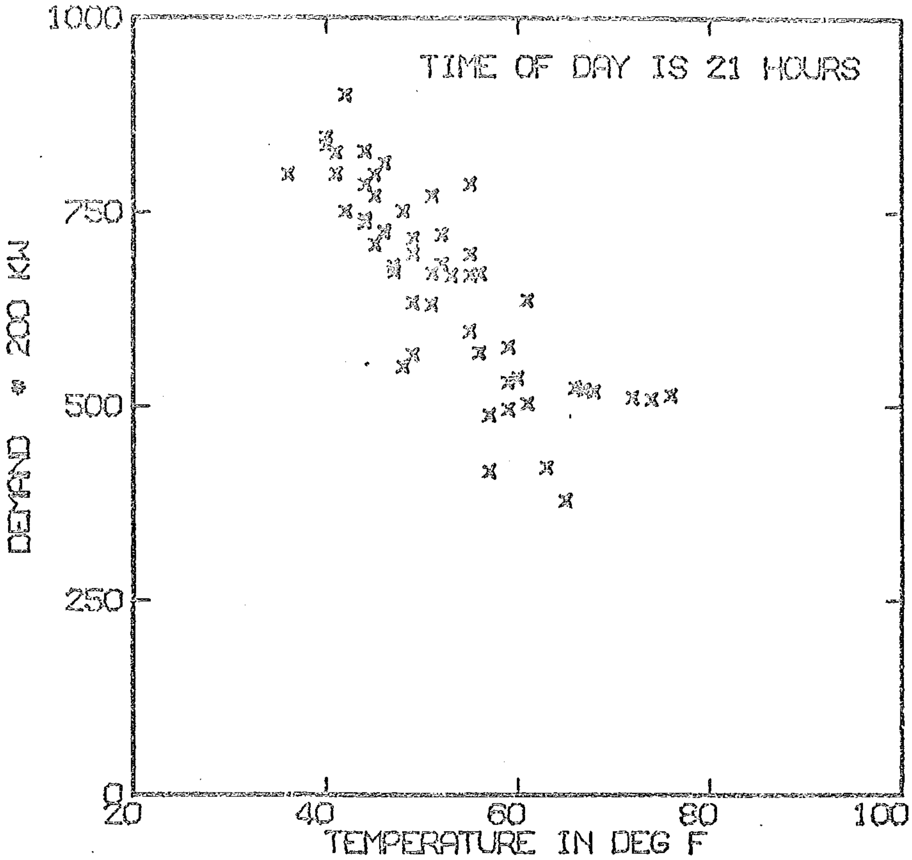

TIME OF DAY IS 16 HOURS
1000
750
500
250
0
20
40
60
80
100
DEMAND * 200 KW
TEMPERATURE IN DEG F
28

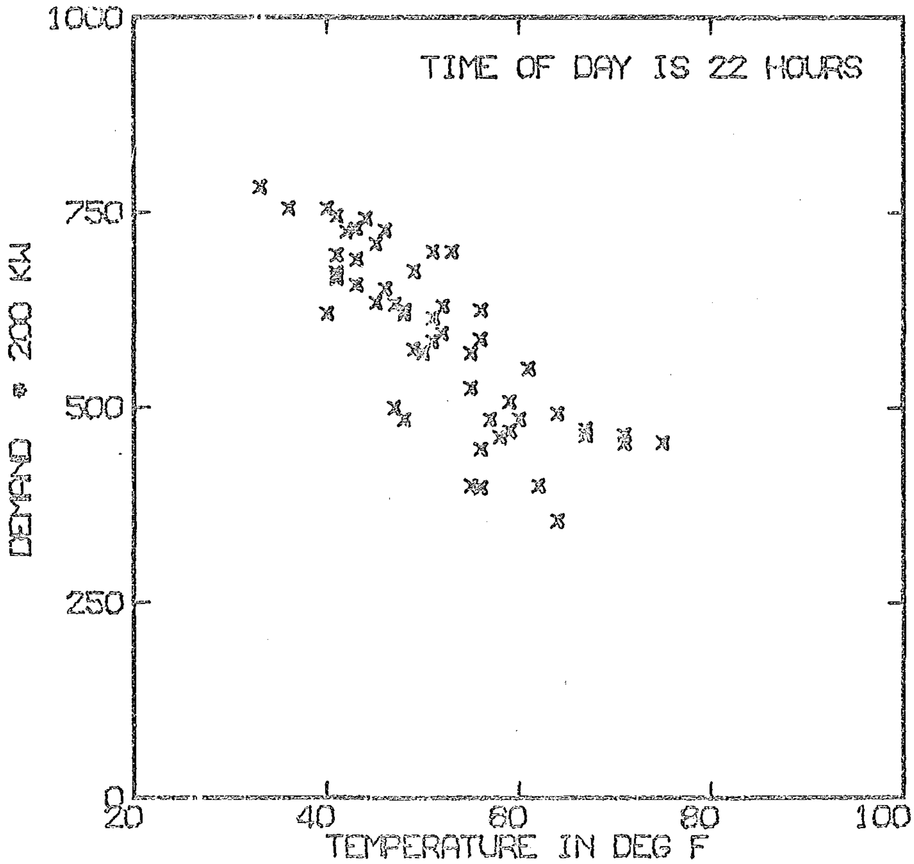

TIME OF DAY IS 17 HOURS
28
1000
750
500
250
0
20
40
60
80
100
TEMPERATURE IN DEG F
DEMAND * 200 KW

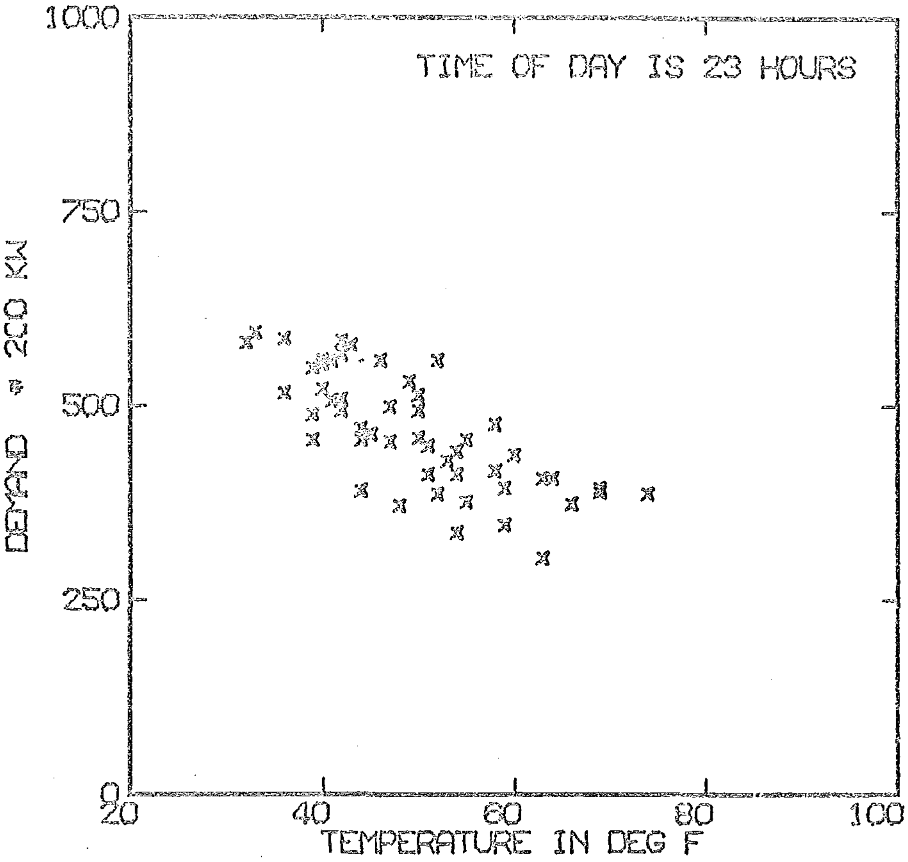

TIME OF DAY IS 18 HOURS
1000
750
500
250
0
20
40
60
80
100
TEMPERATURE IN DEG F
DEMAND * 200 KW

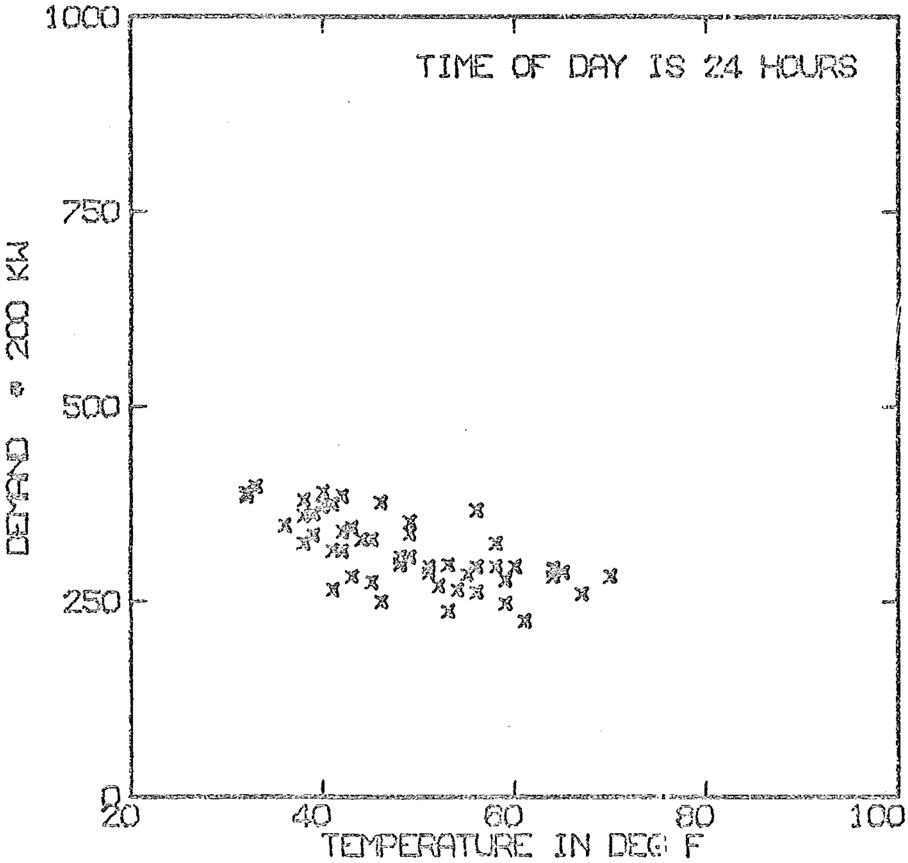

TIME OF DAY IS 19 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
1000
750
500
250
0
20
40
60
80
100


TIME OF DAY IS 20 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
1000
750
500
250
0
20
40
60
80
100


TIME OF DAY IS 21 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
0
250
500
750
1000
20
40
60
80
100


TIME OF DAY IS 22 HOURS
DEMAND * 200 KW
TEMPERATURE IN DEG F
0
250
500
750
1000
20
40
60
80
100


TIME OF DAY IS 23 HOURS
1000
750
500
250
0
20
40
60
80
100
TEMPERATURE IN DEG F
DEMAND * 200 KW


TIME OF DAY IS 24 HOURS
1000
750
500
250
0
20
40
60
80
100
TEMPERATURE IN DEG F
DEMAND * 200 KW

REFERENCES

1. K.N. Stanton and P.C. Gupta, "Forecasting annual or seasonal peak demand in electric utility systems", IEEE Trans. Power Appratus and Systems, vol. PAS-89 No. 5/6, pp 951-959, 1970.

2. P.D. Matthewman and Prof. H. Nicholson, "Techniques for load prediction in the electricity supply industry", Proc. IEE, vol. 115, No. 10, pp 1451-1457, 1968.

3. M.V. Deshpande, "Elements of electrical power station design", Pitman, London, 1966

4. A.R. Cooper, "Load dispatching and the reasons for it, with special reference to the British grid system", Journal IEE, vol. 95, part II, pp 713-732, 1948

5. C. Brewer et. al., "Performance of a predictive automatic load dispatching system", Proc. IEE, vol. 115 No. 10, pp 1577-1586, 1968

6. G.E.P. Box and G.M. Jenkins, "Time series analysis, forecasting and control", Holden Day, San Francisco, 1970

7. Committee to Review Power Requirements, "Report", New Zealand Government Printer, Wellington, 1969

8. Department of Statistics, "New Zealand Year Book, 1970", New Zealand Government Printer, Wellington, 1971.

9. E.B. Mackenzie, "Forecasting and meeting demands for electric power in New Zealand", New Zealand Engineering, vol. 21 No. 11, pp 478-484, 1966

10. W.R. Christiaanse, "Short term forecasting using general exponential smoothing", IEEE Trans. Power Apparatus and Systems, vol. PAS-90 No. 2, pp 900-910, 1971

11. W.R. New, "Load forecasting on the TVA system : Part 1 - substation loads", Trans. AIEE on Power Apparatus and Systems, vol. 81, pp 101-105, 1962

12. W.W. Godard, "Electric utility load forecasting", Trans. AIEE on Power Apparatus and Systems, vol. 74, pp 1428-1440, 1956

13. F.G. Pyatt, "Priority patterns and the demand for household durable goods", University of Cambridge, Department of Applied Economics Monograph 11, Cambridge University Press, 1964

14. M. Davies, "The relationship between weather and electricity demand", Proc. IEE, vol. 106 part C, pp 27-33, 1959

15. Davenport and Root, "Random signals and noise", McGraw - Hill, New York, 1958

16. P.W. Cash and E.C. Scott, "Security of supply in planning and operation of European power systems - part 1", IEEE Trans. Power Apparatus and Systems, vol. PAS-88 no. 1, pp 6-12, 1969

17. R.G. Hooke, "Forecasting the demand for electricity", Trans. AIEE on Power Apparatus and Systems, vol. 74, pp 993-1008, 1955

18. M.J. Turnbull, "Meteorological and electricity demand data", Electrical Engineering Department Memorandum no. 81, University of Canterbury, Christchurch N.Z., 1971

19. R.G. Lipsey, "An introduction to positive economics", Weidenfeld and Nicholson, London, 1968

20. R.G.D. Allen, "Macro-economic theory", Macmillan, London, 1968

21. H.J. Jelinek, "Steady state demand model for power system load prediction", Proc. IFAC Symposium, Institute of Engineers (Australia), Sydney, 1968, pp 99-103

22. G.D. Miller, "The analysis and simulation of electric power system loads", M.E. Thesis, University of Canterbury, Christchurch N.Z., 1968

23. D.J. Bolton, "Electrical engineering economics", Chapman and Hall, London, 1950

24. H.A. Dryar, "The effect of weather on the system load", Trans. AIEE, vol. 63, pp 1006-1013, 1944

25. W.W. Pleines and K.U. Zajic, "A method of estimating demand", Trans. AIEE on Power Apparatus and Systems, vol. 81, pp 879-883, 1963

26. A. Hald, "Statistical theory with engineering applications", John Wiley, New York, 1952

27. J.G. Boggis, "Why bother with load research?", Elect. Rev.(GB), vol. 186 no. 5, pp 176-170, 1970

28. H. Speight, "Economics - the science of prices and incomes", Methuen, London, 1964

29. National Development Conference, "Report of the Fuel and Power Committee", New Zealand Government Printer, Wellington, 1969

30. G.T. Heinemann et. al., "The relationship between summer weather and summer loads - a regression analysis", IEEE Trans. Power Apparatus and Systems, vol. PAS-85 no. 11, pp 1144-1154, 1966

31. J.V.T. Baker, "Forecasting electricity and energy requirements", New Zealand Electrical Journal, vol. 42 no. 9, pp 208-211, 1969

32. R.L. Earle, "Electricity in the food-processing industries", New Zealand Electrical Journal, vol. 42 no. 9, pp 211-214, 1969

33. Department of Statistics, "New Zealand Year Book, 1969", New Zealand Government Printer, Wellington, 1970

34. ——————, "New Zealand Year Book, 1967", New Zealand Government Printer, Wellington, 1968

35. R. Billinton and M.P. Musick, "Spinning reserve criteria in a hydro thermal system by the application of probability mathematics", The Engineering Journal (Canada), vol. 48, pp 40-45, Oct. 1965

36. C.W. Watchorn, "Elements of system capacity requirements", Trans. AIEE, vol. 70 part II, pp 1163-1176, 1951

37. J. Toyoda et.al., "An application of state estimation to short term load forecasting, Part I; Forecasting modelling", IEEE Trans. Power Apparatus and Systems, vol. PAS-89 no. 7, pp 1678-1682, 1970

38. R.A. Hore, "Advanced studies in electrical power system design", Chapman and Hall, London, 1966

39. J.H. Latham et.al., "Probability approach to electric utility load forecasting", IEEE Trans. Power Apparatus and Systems, vol. PAS-87 no. 2, pp 496-504, 1968

40. L.T. Anstine et.al., "Application of probability methods to the determination of spinning reserve requirements for the Pennsylvania - New Jersey - Maryland interconnection", IEEE Trans. Power Apparatus and Systems, vol. 82 no. 10, pp 726-735, 1963.

41. L.L. Garver, "Reserve planning using outage probabilities and load uncertainties", IEEE Trans. Power Apparatus and Systems, vol. PAS-89 no. 4, pp 514-521, 1970

42. A.D. Patton, "Short term reliability calculation", IEEE Trans. Power Apparatus and Systems, vol. PAS-89 no. 4, pp 509-514, 1970

43. R.J. Ringlee et.al., "Frequency and duration methods for power system reliability calculations, Part I; Generation system model", IEEE Trans. Power Apparatus and Systems, vol. PAS-87 no. 10, pp 1787-1796, 1968

44. ———— and A.J. Wood, "Frequency and duration methods for power system reliability calculations, Part II; Demand model and capacity reserve model", IEEE Trans. Power Apparatus and Systems, vol. PAS-88 no. 4, pp 378-388, 1969

45. K. Goldsmith, "The price of reliability in electricity supply", Proc. Conf. on Economics of the reliability of supply, IEE, London, 1967, pp 50-53

46. P.W. Cash and E.C. Scott, "Security of supply in planning and generation of European power systems - part II", IEEE Trans. Power Apparatus and Systems, vol. PAS-88 no. 1, pp 13-21, 1969

47. C.W. Watchorn, "A review of some basic calculations of probability methods as related to power system problems", IEEE Trans. Power Apparatus and Systems, vol. 83 no. 7, pp 737-743, 1964

48. M.P. Musick and C.J. Goodwin, "Availability of hydro-electric generating units in Manitoba", Proc. Conf. on Economics of reliability of supply, IEE, London, 1967, pp 81-85

49. P.W. Cash and F. Faux, "Generation - Central Electricity Generating Board system in England and Wales," Proc. Conf. on Economics of reliability of supply, IEE, London, 1967, pp 1-21

50. P.H. Henault et.al. "Power system long term planning in the presence of uncertainty", IEEE Trans. Power Apparatus and Systems, vol. PAS-89 no. 1, pp 156-164, 1970

51. Planning Committee on Electric Power Development in New Zealand, "Report", New Zealand Government Printer, Wellington, 1969

52. G.O. Wessenauer et.al. "Some problems on a large power system with large generating units", World Power Conference, paper 82 III O/7, Melbourne, 1962, pp 1850-1864

53. R. Billinton and C.T. Goodwin, "Generation capacity planning in Canada", Proc. Conf. on Economics of reliability of supply, IEE, London, 1967

54. ——————— and M.P. Bhavaraju, "Transmission planning using a reliability criterion, Part I ; A reliability criterion", IEEE Trans. Power Apparatus and Systems, vol. PAS-89 no. 1, pp 28-34, 1970

55. M.H. Kent et.al., "Dynamic modelling of loads in stability studies", IEEE Trans. Power Apparatus and Systems, vol. PAS-88 no. 5, pp 756-763, 1969

56. F.C. Ralston, "Predicting central station demand and output", Trans. AIEE, vol. 44, pp 130-136, 1925

57. K.N. Stanton, "Long range demand forecasting for the electric utility industry", Proc. 1969 IEEE PICA Conf. (Denver, Colo.), 1969, pp 565-579.

58. M.J. Turnbull and H.J. Jelinek, "An intermediate range diversified power system demand prediction model", Electrical Engineering Department Memorandum No. 65, University of Canterbury, Christchurch N.Z., 1969

59. I.J. Whiting, "Planning for an expanding electricity supply industry", Op. Research Quarterly, vol. 14 no. 2, pp 107-117, 1963

60. R.L. Wine, "Statistics for scientists and engineers", Prentice-Hall, New Jersey, 1964

61. R.L. Coe, "Opportunity afforded by utility control of consumer loads", IEEE Trans. Industry and General Applications, vol. IGA-5 no. 3, pp 250-254, 1969

62. E.A. Logan, "Is there a limit to the demand for electricity?", Proc. IEE Part A, pp 27-33, 1955

63. R.J. Ball and T. Burns, "An econometric approach to short run analysis of the U.K. economy, 1955-66", Op. Research Quarterly, vol. 19, pp 225-256, 1968

64. R.H. Sarikas and H.B. Thacker, "Distribution system load characteristics and their use in planning and design", Trans. AIEE on Power Apparatus and Systems, vol. 76, pp 564-573, 1957

65. R.F. Hamilton, "Synthetic or equivalent load curves", Trans. AIEE, vol. 61, pp 369-381, 1942

66. W.L. Tadlock, "Evaluation of incompletely diversified loads", Electrical Engineering, vol. 62, pp 485-492, 1943.

67. D.N. Reps, "Monte Carlo simulation of residential transformer loads", IEEE Trans. Power Apparatus and Systems, vol. 82 no. 12, pp 841-864, 1963

68. ASHRAE Guide and Data Book, "Fundamentals and equipment", ASHRAE, New York, 1961

69. W.T. Maunder, "The value of the weather", Methuen, London, 1970

70. R. Freeman, "The accuracy of long range weather forecasts issued by the Meteorological Office", Met. Mag., vol. 95 (1132), pp 321-325

71. G.K.C. Chen and P.R. Winters, "Forecasting peak demand for an electric utility with a hybrid exponential model", Management Science, vol. 12 no. 13, 1966

72. E.D. Farmer, "A method of prediction for non-stationary processes and its application to the problem of load estimation", Proc. 2nd IFAC conf., Basle, 1963, Butterworths, London, 1964, vol. 2, pp 47-54

73. ——————— and M.J. Potton, "Development of on-line load prediction techniques with results from trials in the south west region of the CEGB", Proc. IEEE, vol. 115 no. 10, pp 1549-1558, 1968

74. K. Kawatake and K. Hirasawa, "Adaptive prediction of time series", Electrical Engineering in Japan, vol. 88 no. 8, pp 8-16, 1968

75. M. Toyama and K. Yamada, "Daily load prediction on the basis of Farmer's Method", Electrical Engineering in Japan, vol. 87 no. 7, pp 50-58, 1967

76. W.J. Maunder, "Temperature Forecasts and the assessment of electric power demand in New Zealand", New Zealand Met. Service, Technical Note 195, April 1971

77. C. Baldwin et.al. "Load and capacity models for generation planning by simulation", Trans. AIEE on Power Apparatus and Systems, vol. 74, pp 993 -1008 1955

78. The Electricity Council, "Annual report and accounts, 1964-1965", HMSO, London, 1965

79. F.M. Fisher, "The identification problem in econometrics", McGraw-Hill, New York, 1966

80. T.C. Angus, "The control of indoor climate", Pergamon, London, 1968

81. L. Bastings and Simmons, "Climatic zones and domestic heating in New Zealand", New Zealand J. Sci, Tech., vol. 32 no. 3, pp 44-53, 1945

82. H.H. Rosenbrock, "An efficient method for finding the greatest or least value of a function", Computer Journal, vol. 3, pp 175-184, 1960

83. K. Fukunaga and W.L.G. Koontz, "Application of the Karhuren-Loeve expansion to feature selection and ordering", IEEE Trans. Computers, vol. C-19 no. 4, pp 311-318, 1970

84. L. Bastings, "Insulation and heating of buildings", Information Series no. 18, New Zealand Dept. Scientific and Industrial Research, 1964

85. R.E. Lacy, "Variations of the winter means of temperature, wind speed and sunshine, and their effects on the heating requirements of a house", Meteorological Magazine, vol 80, pp 161-166, 1951

86. L. Bastings, "Sunshine and shade. A handbook for New Zealand", Building Research Bureau of New Zealand Inc. 1967

87. D.K.A. Gillies et.al. "A new approach to forecasting daily peak loads", Trans. AIEE on Power Apparatus and Systems, vol. 75, pp 382-387, 1958

88. S.R. Searle and J.G. Udell, "The use of regression on dummy variables in management research", Management Science, Series B, vol. 16 no. 6, pp B-397-409, 1970.

89. R.J. and T.H. Wonnacott, "Econometrics", John Wiley, Toronto, 1970.

90. J. Shiskin, "Decomposition of economic time series", Science, vol. 128 no. 3338, pp 1539-1546, 1958

91. R.G. Brown, "Smoothing, forecasting and prediction of discrete time series", Prentice-Hall, New Jersey, 1963

92. H.M. Runyan, "Cybernetics of economic systems", IEEE Trans. Systems, Man and Cybernetics, vol. SMC-1 no. 1, pp 8-18, 1971

93. H.G. Berrisford, "The relation between gas demand and temperature : A study in statistical demand forecasting", Op. Research Quarterly, vol. 16 no. 2, pp 229-246, 1965

94. D.O. Sewell, "Electric household durable goods - Economic aspects of their manufacture in New Zealand", New Zealand Institute of Economic Research, Research paper No. 7, Wellington N.Z.,1965

95. R.J. and T.H. Wonnacott, "Econometrics", John Wiley, New York, 1970

96. F.M. Fisher and C. Kaysen, "A study in econometrics: The demand for electricity in the United States", North-Holland, Amsterdam, 1962

97. L. Higgens and S. Brubacher, "A new method of forecasting hourly demands", Eighth World Energy Conference, Bucharest, paper 4.1-92, June 1971.

98. S.N. Marris, "A technical survey of problems and methods of seasonal adjustment in Europe and North America with special reference to the United States Bureau of the Census Method II", Proc. Conf. on Seasonal Adjustment on Electronic Computers, OECD, Paris 1970, pp 35-78.

99. S.D. Goldring and E.N. Zesch, "Urban residential load survey", Union Electric Co., Oct. 1962 (not published).